



UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE ARTES E COMUNICAÇÃO
DEPARTAMENTO DE LETRAS
PROGRAMA DE PÓS-GRADUAÇÃO EM LETRAS

ERICO MONTEIRO DA SILVA

**A NATUREZA E A ORDEM DOS ADVÉRBIOS PARA ANÁLISES E DESCRIÇÕES
AUTOMÁTICAS**

Recife

2018

ERICO MONTEIRO DA SILVA

**A NATUREZA E A ORDEM DOS ADVÉRBIOS PARA ANÁLISES E DESCRIÇÕES
AUTOMÁTICAS**

Dissertação apresentada ao Programa de Pós-Graduação em Letras da Universidade Federal de Pernambuco, como requisito parcial à obtenção do Título de Mestre em Letras.

Área de concentração: Linguística

Orientador: Prof. Dr. Marcelo Amorim Sibaldo

Recife

2018

Catálogo na fonte
Bibliotecária Jéssica Pereira de Oliveira, CRB-4/2223

S586n Silva, Erico Monteiro da
A natureza e a ordem dos advérbios para análises e descrições
automáticas / Erico Monteiro da Silva. – Recife, 2018.
155f.: il.

Orientador: Marcelo Amorim Sibaldo.
Dissertação (Mestrado) – Universidade Federal de Pernambuco. Centro
de Artes e Comunicação. Programa de Pós-Graduação em Letras, 2018.

Inclui referências.

1. Advérbio. 2. Gramática Gerativa. 3. Gramática Livre de Contexto.
4. Parser. 5. Python-NLTK. I. Sibaldo, Marcelo Amorim (Orientador).
II. Título.

410 CDD (22. ed.)

UFPE (CAC 2020-85)

ERICO MONTEIRO DA SILVA

**A NATUREZA E A ORDEM DOS ADVÉRBIOS PARA ANÁLISES E DESCRIÇÕES
AUTOMÁTICAS**

Dissertação apresentada ao Programa de Pós-Graduação em Letras da Universidade Federal de Pernambuco, como requisito parcial à obtenção do Título de Mestre em Letras.

Aprovada em: 01/03/2018.

BANCA EXAMINADORA

Prof. Dr. Marcelo Amorim Sibaldo (Orientador)

Universidade Federal de Pernambuco

Prof. Dr. Rafael Bezerra de Lima (Examinador Externo)

Universidade Federal Rural de Pernambuco

Prof. Dr. Adeilson Pinheiro Sedrins (Examinador Externo)

Universidade Federal Rural de Pernambuco

AGRADECIMENTOS

À Inteligência suprema e Criadora de todas as coisas, Deus.

Ao meu orientador Prof. Dr. Marcelo Amorim Sibaldo, por ter aceito este desafio e com enorme paciência e orientação me guiou na elaboração e concretização desta dissertação.

Ao Prof. Dr. Rafael Bezerra Lima por suas considerações, tanto na qualificação quanto na defesa. Agradeço por sempre estar me guiando nos meus mais importantes afazeres acadêmicos.

Ao Prof. Dr. Adeilson Pinheiro Sedrins por ter aceito participar da banca e pelas grandes considerações.

Verônica, a quem agradeço por sempre estar ao meu lado. Com você, o fenômeno de dois universos que insistem em colidir e se unir, se unir e colidir, formando assim a busca da verdade e do amor acontece, todos os dias.

A minha família que deposito amor, afeto e confiança, em especial, a minha mãe, Glorineide e, a meu pai, Edvaldo, que nesta vida vieram e aceitaram a missão onde os laços profundos se enraízam na alma e duram por toda a eternidade, meu amor e gratidão.

A Eduardo Órácio, amigo que compartilhei vida durante este tempo acadêmico, obrigado pelo aprendizado e companheirismo. Crescemos juntos neste período.

A Josiene Melo, pelo companheirismo e crescimento mútuo neste tempo de desenvolvimento de vida e acadêmico.

Aos amigos Anderson Almeida, Edney Alexandre Belo e Sivaldo Correia, amigos que encontrei e me acompanharam neste percurso acadêmico e além, os meus sentimentos verdadeiros.

Aos colegas e amigos do Getegra-UFPE, em especial, Alane Luma e Déreck Pereira, obrigado por tornar possíveis as discussões e ampliação dos conhecimentos (gerativos) adquiridos e massificados.

Aos professores Dr^a. Claudia Roberta Tavarez Silva, Dr. Antônio Xavier, Dr. Vicente Macip e demais professores que direta e indiretamente contribuíram na minha formação e desenvolvimento nesta etapa acadêmica.

Aos demais colegas de curso que estimo e torço pelo sucesso nesta estadia terrena, na vida acadêmica e além.

“Poemas são feitos por tolos como eu, mas só Deus pode criar uma árvore; e só Deus, que cria as árvores, cria tolos como eu. Mas, sabe, só os tolos como eu podem criar um Deus que cria uma árvore.” (HARBURG, 1920, p. 7)

RESUMO

Esta dissertação tem por objetivo a busca da representação e decodificação dos advérbios a partir da sua natureza e traços para a implementação de sistemas de processamento da linguagem natural e o desenvolvimento de analisadores sintáticos denominados *parsers*. Para tanto, a pesquisa segue o método proposto por Dias da Silva (1996), método aplicado no desenvolvimento e ordem estrutural do trabalho, em que utiliza a sequência linear de fases nos estudos e pesquisas que tratam, por meio de computadores, estruturar por fases a análise do fenômeno linguístico a ser modelado. Essas fases compreendem a fase Linguística, que visa a descrição e análise do objeto com o objetivo de resgatar o percurso histórico e descritivo da categoria nos estudos gramaticais de vertente grega, tradicional e linguística de cunho gerativo; a fase Linguístico Computacional, a qual visa a formalização teórica e do objeto a partir do processamento da Linguagem Natural e as descrições dos modelos formais; a fase Computacional, que codifica a modelagem para o uso de fato nos sistemas de processamento da linguagem natural partindo da gramática livre de contexto e dos traços unificados na estrutura arbórea para, assim, partir para a modelagem e codificação dos resultados parciais por meio da linguagem de programação Python e da biblioteca NLTK. A pesquisa demonstrou a modelagem da categoria tanto nos processos linguísticos quanto computacionais na busca dos traços que licenciam a natureza e a ordem dos advérbios em um empreendimento que modele os advérbios do tipo Altos e Baixos vistos como uma possibilidade de organização e análise para uso em sistemas que simulem a competência humana.

Palavras-chave: Advérbio. Gramática Gerativa. Gramática Livre de Contexto. Parser. Python-NLTK.

ABSTRACT

This dissertation aims to search representation and decoding of adverbs from its nature and features for the implementation of natural language processing systems and the development of parsers called parsers. Therefore, the search follows the method proposed by Dias da Silva (1996), in the development method used and structural work order, in which it uses the linear sequence of phases in the studies and researches that deal, through computers for to structure the analysis of the linguistic phenomenon to be modeled. These phases comprise the Linguistic phase, which aims at describing and analyzing the object with the purpose of retrieving the historical and descriptive course of the category in the grammatical studies of Greek, traditional and linguistic aspects of a generative nature; the Computational Linguistic Phase, which aims at the theoretical and object formalization from the processing of the Natural Language and the descriptions of the formal models; the Computational phase, which encodes the modeling for the use of fact in the natural language processing systems starting from the context-free grammar and the unified features in the tree structure, to begin with the modeling and coding of the partial results through the language the Python programming language and the NLTK library. The research has demonstrated the modeling of the category both in linguistic and computational processes in the search for traits that license the nature and order of adverbs in an enterprise that models the adverbs of the type Low and High seen as a possibility of organization and analysis for use in systems that simulate human competence.

Keywords: Adverb. Gerative Grammar. Context-Free Grammar. Parser. Python-NLTK.

LISTA DE ABREVIATURAS E SIGLAS

AA	Adjunto Adverbial
AC	Adjunto Circunstancial
Adv	Advérbio
Ant	Anteposição
AO	Adjunto Oracional
Asp	Aspecto
Aux	Verbo auxiliar
Atr	Atributo
CF	Categoria Funcional
CFG	Gramática Livre de Contexto (Context-Free Grammar)
D	Determinante
FF	Forma-Fonética
FL	Forma-Lógica
G	Gramática (Formal)
Gen	Gênero
GG	Gramática Gerativa
GT	Gramática Tradicional
Int	Intensificador
INFL	Flexão (Inflection)
L	Língua (Formal)
LC	Linguística Computacional
LCA	Axioma de Correspondência Linear (ACL) - Linear Correspondence Axiom
LFG	Gramática Léxico-Funcional (Lexical-Functional Grammar)
MAV	Matrizes de Atributo-Valor
N	Nome
Num	Número
PB	Português Brasileiro
PE	Português Europeu
Per	Pessoa
PM	Programa Minimalista
PNL	Processamento Natural da Linguagem
P&P	Princípios e Parâmetros

AP	Sintagma Adjetivo
AdvP	Sintagma Advérbio
CP	Sintagma Complementizador
DP	Sintagma de Determinante
NP	Sintagma Nominal
TP	Sintagma de Tempo
VP	Sintagma Verbal
T	Tempo
TRL	Teoria da Regência e Ligação
V	Verbo
v	Verbo leve (vezinho)
e	Conjunto vazio de uma L
>	Preceder

SUMÁRIO

1	INTRODUÇÃO.....	12
2	OS ESTUDOS DESCRITIVOS E CIENTÍFICOS DO ADVÉRBIO.....	18
2.1	INTRODUÇÃO.....	18
2.1.1	Os primeiros estudos – a vertente grega.....	19
2.1.2	Nas gramáticas tradicionais.....	22
2.2	QUESTÕES DESCRITIVAS.....	38
2.3	O ADVÉRBIO NOS ESTUDOS DA GRAMÁTICA GERATIVA.....	42
2.3.1	Advérbios como Adjuntos.....	43
2.3.2	Jackendoff (1972).....	48
2.3.3	Pollock (1989).....	51
2.3.4	Laenzlinger (1998).....	53
2.3.5	Cinque, 1999, 2004, 2006.....	57
2.4	A DELIMITAÇÃO DOS ADVÉRBIOS.....	63
2.4.1	A ordem dos advérbios.....	64
2.4.2	Advérbios Altos.....	67
2.4.3	Advérbios Baixos.....	74
2.5	CONSIDERAÇÕES.....	79
3	O LINGÜÍSTICO E O LINGÜÍSTICO COMPUTACIONAL.....	82
3.1	O PROGRAMA MINIMALISTA.....	89
3.2	MODELOS FORMAIS.....	94
3.2.1	O parser, a gramática e o léxico.....	107
3.2.2	O Parser.....	108
3.2.3	A Gramática.....	112
3.2.4	A linguagem Python e o NLTK.....	119
3.3	CONSIDERAÇÕES.....	126
4	ANÁLISE E MODELAÇÃO COMPUTACIONAL.....	128
4.1	MODELAÇÃO E ANÁLISE.....	128
4.2	O <i>PARSER</i> E AS ANÁLISES POSSÍVEIS.....	138
4.3	CONSIDERAÇÕES.....	143
5	CONSIDERAÇÕES FINAIS.....	146
	REFERÊNCIAS.....	149

1 INTRODUÇÃO

As subáreas de conhecimento na Linguística, em que os nomes são compostos, normalmente abarcam as questões teóricas e empíricas de forma combinadas. A Linguística Computacional combina o conhecimento da área da Linguística e da Computação, assim como, por exemplo, a Sociolinguística que utiliza a Linguística e a Sociologia para o desenvolvimento das pesquisas sobre variações linguísticas existentes nos diversos extratos sociais em uma dada língua natural. Neste sentido, a combinação da Sintaxe com outra subárea de investigação da Computação conduz a um campo de investigação mais específico, no caso, os estudos que abordem a competência linguística humana no que tange as análises sintáticas, ou seja, o sistema computacional das línguas naturais visando uma emulação da capacidade humana de construir e reconhecer sentenças pertencentes a língua alvo.

Combinando tais esferas do conhecimento, a Sintaxe e a Computação, deriva uma forma de abordagem para a análise sintática conhecida como parser. O parser na esfera linguística apresenta duas dimensões teóricas e práticas, a primeira é representacional, ou seja, um modelo de representação das estruturas sintáticas (constituída por traços funcionais e lexicais, princípios universais e parâmetros comuns ou particulares, dentre outros), e a segunda como um elemento psicologicamente motivado que permite ao ser humano interpretar as sentenças. Na esfera computacional, o *parser* é compreendido como uma análise sintática (em inglês denominado *parsing*) em que o processo de análise de sequências de entrada (*input*) determina a estrutura de uma dada gramática formal. Vale ressaltar que o *parser* é constituído por um dicionário (itens lexicais e semânticos) e regras (definidas por uma Gramática-Livre de Contexto) para estruturar os dados e organizá-los em estruturas arbóreas, de linhas e colunas, em matrizes de atributo valor, ou outros modelos representacionais existentes.

Diante dessas duas possibilidades, o foco principal da pesquisa deverá dar um direcionamento mais Linguístico ou mais Computacional, ou seja, se o foco for a formalização e modelação dos componentes da língua e dos fenômenos ligados a linguagem, as questões da Sintaxe serão mais relevantes que a eficiência computacional, caso seja o foco as questões computacionais, a consistência das explicações linguísticas serão irrelevantes diante dos resultados computacionais obtidos na implementação dos modelos linguísticos no computador. Unir tais possibilidades consiste no desafio de alinhar as questões teóricas e empíricas à modelagem computacional e implementacional os fenômenos a que se deseja, sejam elas fonéticas, morfológicas, sintáticas, semânticas e pragmáticas.

Tratar de evidenciar as propriedades sintáticas de um dos componentes da língua que desencadeiam fenômenos linguísticos, neste caso, o advérbio, no campo das pesquisas formais em Linguística e do Processamento Natural da Linguagem (PNL) para o uso e desenvolvimento dos analisadores sintáticos automáticos, o já referenciado *parser*, se mostra um desafio perseguido por pesquisadores da ciência da linguagem, incluindo os que se detêm nos estudos da Linguística Computacional. Com a finalidade de dotar com maior robustez e eficiência os *parsers*, partindo deste fenômeno da língua categorizado como advérbio, é observado uma necessidade de pesquisa exaustiva devido a sua natureza heterogênea que dificulta as análises linguísticas quando necessita conceituar precisamente esta categoria, somando a isto, há as dificuldades na estruturação das regras utilizadas nos *parsers* devido a ordem dos advérbios variarem dificultando as regras mais deterministas. Na tentativa de abarcar esta categoria, assumimos alguns conceitos teóricos da perspectiva Cartográfica em relação aos advérbios, que conceitua-os em Altos e Baixos.

Ao analisar sentenças com itens pertencendo a categoria advérbio de forma mais robusta, pretende-se capacitar o mais próximo possível da capacidade humana de formalizar e descrever uma sentença e estruturá-la de acordo com pressupostos linguísticos, e isto é o mínimo para uma ferramenta que propõe a ser uma ferramenta computacional no auxílio das pesquisas linguísticas. A busca por evidências dos traços sintáticos dos advérbios que os licenciem nas suas posições é uma constante nos estudos linguísticos, tendo em vista que nem todos os advérbios possuem mobilidade “livre” devido a restrições. Também, os advérbios se encontram diante de impasses quanto a sua natureza, seja ela um item categorial lexical ou uma subclasse dos adjetivos, torna a pesquisa envolvendo a construção de um modelo linguístico computacional, a partir dos resultados obtidos dos estudos linguísticos sobre o tema, desafiador. Este empreendimento preenche uma lacuna em relação a essas pesquisas que utilizam *parser*, tanto na utilização, quanto na criação, encontram entraves ao utilizar essa “conturbada” categoria.

Nas gramáticas tradicionais, os advérbios são definidos como palavras que essencialmente se agregam e modificam o verbo para expressar, por exemplo, as circunstâncias ou intensificam a qualidade que o adjetivo atribui ao substantivo. Miotto, Silva e Lopes (2013) evidenciam com o exemplo da palavra “provavelmente”, definida como advérbio pela gramática tradicional, se posiciona em diferentes lugares na sentença, graças a essa mobilidade, possibilita modificar categorias diversas, além das canônicas (verbo, adjetivo e advérbio). As questões pertinentes em torno dessa categoria heterogênea centra-se na busca das razões que fazem o advérbio figurar como uma categoria independente, já que há traços

que possibilitam a natureza sintática e semântica a torná-la uma categoria lexical independente. As questões se estendem em relação aos traços e a natureza dos itens classificados como tais, possibilitando formalizações linguísticas e linguísticos computacionais.

A descrição que pretendemos visa uma implementação linguístico computacional sobre os traços sintáticos gerais dos advérbios e sua ordem para explicitá-los computacionalmente, para isso, um levantamento de alguns estudos gramaticais já realizados é imprescindível, partindo dos gramáticos tradicionais até importantes gerativistas que versaram sobre o tema com diversas propostas. Seguimos o gerativismo, na fase do Programa Minimalista (PM), proposto por Chomsky (1995) na estruturação das regras sintáticas, que objetiva organizar e diminuir o aparato técnico da Teoria de Princípios e Parâmetros. Tal empreendimento traz uma elaboração do Sistema Computacional que poderá nos fornecer mais informações de licenciamento dos traços existentes e seguindo alguns princípios que determinam os valores dos advérbios na organização do conjunto de uma estrutura lexical e funcional que detalha sua natureza, a perspectiva Cartográfica, desenvolvida por Rizzi (1997) e, utilizado por Cinque (1999) na busca da natureza do advérbio e, aplicado no Português Brasileiro (PB) por Sant'ana (2010). É considerado uma das teorias de cunho gerativo que detalha minuciosamente os itens lexicais e suas projeções funcionais.

Tendo em vista que os principais analisadores automáticos do PB analisam as sentenças do português brasileiro, não se detêm sobre as regras sintáticas dos advérbios com mais detalhes, tampouco fazem um maior uso dos trabalhos linguísticos que tratam dessa classe, apenas o localizam na estrutura de constituintes e limitando-os a modificadores e circunstâncias do verbo. Diante disso, a justificativa de catalogar os estudos acerca dos advérbios se faz pertinente, tanto para evidenciar a necessidade de uma maior abrangência a respeito do advérbio, quanto para propor formas de análises computacionais mais robustas a respeito dessa classe.

Uma das opções para explorar os resultados obtidos da formalização é explicitá-los e configurá-los com a unificação dos traços sintáticos dos advérbios, já que, até o momento não ganhou atenção na literatura gerativa, proposta por Chomsky (1965, 1986, 1995, 2001) e cia, pois os advérbios não foram formalizados exaustivamente, muito por conta da falta de consenso em torno do advérbio e das expressões adverbiais, somado a uma melhor descrição e análise profunda da relação deste item com o Sistema Computacional Mental.

Sobre a formalização de um *parser*, é exigido alguns mecanismos computacionais como a utilização de uma linguagem de programação que possa viabilizar o empreendimento,

neste quesito é recomendado o uso da linguagem de programação Python(3) em conjunto da sua biblioteca destinada ao uso do processamento da linguagem natural, a *Natural Language ToolKit* (conhecida como NLTK) (BIRD; KLEIN; LOPER, 2014). A utilização dos traços unificados na formalização é justificado por ser um modelo estruturado que torna visível qualquer traço explícito na própria projeção, favorecendo não só a descrição linguística, mas também a análise linguístico computacional. Por facilitar uma estrutura mais lexicalista e semanticamente explicativa em nível linguístico, o NLTK possui módulos que operam nesse nível descritivo, favorecendo as possibilidades de análises do *parser* nas expressões linguísticas fornecidas pelos usuários.

Muitas áreas do conhecimento investem em programas para facilitar tarefas de pesquisa, praticas e uso dos dados recolhidos, como a Arquitetura, a Medicina, o Direito, dentre outros campos de pesquisa. Principalmente quando a tarefa utiliza uma gama maior de dados ou modelagens mais rigorosas sobre uma aplicação dessas áreas. Na área da Linguística, o uso dos softwares já são uma realidade, contudo, ainda não possuem abrangência de uso em determinados campos da Linguística, muito por falta de interesse dos pesquisadores em fornecer os modelos que os softwares necessitam para auxiliarem os empreendimentos mais rígidos. Por outro lado há no mercado etiquetadores, manipuladores textuais, *corporas* eletrônicos, assistentes virtuais etc., que são realidades disponíveis, mas pouco utilizados por linguistas.

Em relação aos analisadores sintáticos disponíveis para o PB, a tendência da maioria é a análise à nível de constituintes das sentenças, respeitando a ordem horizontal e linear da língua (fala, leitura e escrita, da esquerda para a direita) e estruturando a árvore em sentido vertical de cima para baixo. Por razões de limitação das ferramentas ou pela complexidade linguística, acabam analisando de maneira geral e simples as sentenças dos usuários sem se atentar ou marcar os traços gramaticais ou sintáticos dos itens localizados na estrutura. Há trabalhos de linguistas que se preocupam em implementar partes específicas do sistema computacional, como o trabalho de Santos (2009) sobre o movimento do verbo e sua projeção ao IP; os trabalhos de David (2007), David e Othero (2011) sobre os agrupamentos nominais se configuram como bons avanços nessas pesquisas mais computacionais. Apesar dos trabalhos observados servirem de apoio aos pesquisadores que se debruçam sobre o tema, os modelos desenvolvidos são mais restritos a outras categorias e não se detêm no advérbio, por limitação do objeto ou por não darem conta da problemática que esta categoria dispõe em qualquer trabalho, deixam a cargo de trabalhos futuros de outros pesquisadores.

No nível computacional, trabalhos que elaboram regras computacionais, modelando a teoria linguística para uso formal, são comuns nesses tipos de pesquisa, ainda assim, há poucas pesquisas que versam sobre o minimalismo na criação de *parser* que modelem fenômenos mentais e o Sistema Computacional de forma mais abrangente. Tentativas de formalizações específicas, como o trabalho de Veenstra (1998), proporcionam fôlego ao tema, entretanto, outros trabalhos de cunho gerativo são realizados em larga escala, utilizando modelos alternativos que diferem dos modelos transformacionais chomskyanos, em que há um maior número de traços e são estruturados por Matrizes de Atributo-Valor (MAV), a exemplo de Bresnan e Kaplan (1982) e Falk (2001) com os modelos da LFG (*Lexical-functional grammar*) e de Pollard e Sag (1994), com a HPSG (*Head-driven phrase structure grammar*), tais modelos são difundidos com facilidade quando se tratam da modelação computacional de cunho gerativo.

Como o objetivo é a formalização de um *parser* que analise sentenças com advérbios de acordo com os pressupostos da teoria gerativa, tanto na fase atual do Programa Minimalista (PM) quanto as características da categoria de acordo com as hipóteses da teoria Cartográfica, um conjunto de perspectivas descritivas e teóricas sobre os advérbios foram necessárias para um resultado mais adequado, em vista que um *parser* tende a analisar de forma geral as sentenças. Tentamos abarcar os advérbios no conceito, nas descrições e nas análises que essa categoria foi e é submetida, desde a antiguidade com os gregos que definiram as características mais perceptíveis e os gramáticos tradicionais que se debruçaram sobre a categoria organizando a classificação, até as pesquisas linguísticas que possuem um poder de análise e descrição mais apurados com aporte científico que podem ser aplicados em analisadores automáticos computacionais destinados à análise de sentenças do português brasileiro. Por conta da abrangência do tema, tanto linguístico quanto linguístico computacional, após a exposição geral da categoria, traçamos os caminhos determinados por Cinque (1999) para a possibilidade de definir a ordem e a natureza dos itens classificados como advérbios e aplicá-los na formalização computacional resultante das conclusões da pesquisa. Tendo determinado os limites de abrangência, a análise se deterá nos advérbios ditos Altos e Baixos.

A organização da dissertação segue uma divisão que se assemelha a sequência metodológica delineada por Dias da Silva (1996) sobre os estudos e pesquisas que tratam dos fenômenos linguísticos e o uso de computadores. Seguindo a proposta, no primeiro domínio linguístico, o capítulo intitulado “Os estudos descritivos e científicos do Advérbio” discute os estudos acerca dos advérbios. Partindo dos estudos gregos devido a sua importância

relacionado as primeiras descrições morfológicas e sintáticas e a busca da natureza deste item no discurso (fala). Tal início justifica a inserção da gramática tradicional, por ser o primeiro campo a absorver e tentar solucionar as problemáticas advindas desta categoria. Por fim, os estudos científicos capitaneados na área linguística pelos gerativistas, a fim de utilizarmos o patrimônio teórico-analítico sobre os fenômenos linguísticos para, assim, delinear os dados a serem analisados posteriormente. O capítulo, intitulado “O Linguístico e o Linguístico Computacional”, continua inicialmente o domínio linguístico tratando da Teoria Gerativa moderna, em especial, o Programa Minimalista, para então seguir o próximo domínio, o Linguístico Computacional, no qual as Representações e formalizações são abordadas e aplicadas no domínio Computacional, introduzindo alguns conceitos para a codificação das representações nos formalismos e aplicações da linguagem de programação Python-NLTK.

O quarto capítulo sistematiza o conteúdo dos dados e das representações que são utilizadas na análise e descrição sintática dos advérbios especificados como Altos e Baixos, para a implementação de um *parser*, demonstrando as possibilidades e limitações da modelação sobre a natureza e tipo dos advérbios, além da ordem que eles podem ocupar, para assim, termos uma base futura para a resolução da formalização e modelação linguística do PM e conseqüentemente possibilitar de forma detalhada os advérbios analisados por um *parser* mais robusto.

2 OS ESTUDOS DESCRITIVOS E CIENTÍFICOS DO ADVÉRBIO

2.1 INTRODUÇÃO

Em vista da sua presença e tradição nos estudos das línguas naturais, iniciamos nossas considerações a respeito dos advérbios a partir da Gramática Tradicional (GT). A expressão Gramática Tradicional¹ aqui utilizada se refere ao caráter historiográfico e descritivo, tal sentido, de acordo com Azevedo (2015), foca na evolução do pensamento gramatical, cobrindo um grande conjunto de conhecimentos sobre a estrutura e uso das línguas, um estudo que vem desde a Antiguidade grega e latina até a metade do século XIX. As definições dos advérbios na GT servem como ponto de partida à problemática causada por esses itens na categorização, graças a sua heterogeneidade lexical e gramatical, além das inúmeras posições que determinados advérbios assumem na sentença.

Vale destacar que a GT recobre um amplo conjunto de conhecimentos sobre a estrutura e uso da língua, objetivando o domínio e emprego das formas “corretas” da linguagem falada e escrita. É preciso salientar que a GT possui fundamentos históricos e por vezes, a não objetividade diante da gramática, vista como algo pronto, paira no imaginário popular. Mas até que ponto os estudos presentes nas gramáticas servem à linguística? Ou os conceitos sobre o advérbio servem para os estudos linguísticos? Veremos que até a GT sofre com essa classe heterogênea.

Antes de adentrarmos nas gramáticas tradicionais do PB e nas suas considerações sobre o advérbio, buscaremos o legado que a gramática tradicional herda na sua tradição, a partir da linguista e pesquisadora Neves (2005) que traz fontes e estudos anteriores da gramática grega. Selecionamos apenas os estudos gregos, sem adentrarmos posteriormente nos estudos latinos por este ser influenciado no grego, segundo Neves (2005, p. 642), “a existência de um legado grego ao modo de organização da gramática portuguesa, como representante que é da gramática ocidental, oriunda da latina, que é calcada na grega” nos dá a percepção de que a gênese do nosso objeto se encontra com os gregos e se perpetua na latina.

Posteriormente à observação do legado grego, adentramos nas definições dos advérbios nas GT que atribuem uma função semântica de modificador do verbo, do nome, do adjetivo e do próprio advérbio, considerando as características circunstanciais e modificadoras

1. 1 A Gramática Tradicional direcionada as atividades pedagógicas voltadas ao ensino de leitura e escrita como requisito para o domínio das formas corretas da língua falada e escrita não serão abordadas, tendo em vista que essa perspectiva possui algumas problemáticas no campo do ensino e aprendizado, questões que fogem do objetivo da pesquisa.

exclusivamente ligadas às questões semânticas e sintáticas, como veremos nesse percurso histórico. A justificativa em adentrar nessa questão primeiro no legado grego se dá pelo fato historiográfico descritivo ser datado a partir das observações de Platão e Aristóteles a respeito da língua datarem na Antiguidade grega, por isso passaremos primeiro pelo legado grego a respeito dos advérbios, já que buscamos algumas evidências sobre nosso problema nas fontes primeiras da GT. Em relação a GT da língua portuguesa, iniciaremos por Fernão de Oliveira (1536), considerado um dos primeiros gramáticos a dar início a gramática portuguesa; posteriormente pelo pensamento filosófico que pairava no século XVII e XVIII com a gramática filosófica de Barbosa (1866), que especulava a língua como expressões mentais universais na espécie humana; a sistematização de Oiticica (1955); a gramática portuguesa mais independente com Barros (1540); os estudos gramaticais de Bechara (2009) além de outros gramáticos.

2.1.1 Os primeiros estudos – a vertente grega

De acordo com Neves (2005), a partir de toda uma tradição em torno dos estudos sobre a linguagem e a língua, vistas em Platão (427 a.C.-347) e Aristóteles (384 a.C.-322 a.C.), e posteriormente pelos estoicos no período helenístico, Aristarco é considerado o primeiro gramático a identificar oito partes do discurso (nome, verbo, particípio, pronome, artigo, advérbio, preposição e conjunção). Dionísio o Trácio, um alexandrino seguidor da escola Aristarco, foi o organizador da arte da gramática com grande influência que ainda hoje perdura em muitas obras gramaticais do Ocidente. Sua obra dá destaque à analogia dos sons das palavras e à morfologia.

A delimitação e método da gramática ganham forma com outro alexandrino, Apolônio Díscolo, que na metade do século II a.C sistematiza todos os estudos linguísticos sobre uma base filosófica, ou seja, sua obra se fundamenta em uma filosofia da linguagem e os princípios da sua filosofia possuíam características universais, além de discorrer sobre a Sintaxe. Neves (2005) nos indica que a partir desses gramáticos, uma tradição se fundamenta e influencia até hoje as gramáticas ocidentais.

Nos primeiros estudos sobre os advérbios, Neves (2005) nos mostra que Aristóteles os colocava no grupo dos Nomes, mas o seu papel na sentença e a sua colocação na frase o levaria a se distanciar dos Nomes. Neves (2005) também aponta que o advérbio, já como classe em Dionísio, recebe o nome *mesótes* (intermédio) pelo fato de não possuir o caso de desinência de gênero, correspondente ao genitivo plural. O advérbio é considerado, graças a

sua função, um tipo de adjetivo do verbo e recebe o nome de *epírrhema*, corresponde ao termo *epítheton* (adjetivo) para a classe do Nome. Os estoicos sugeriram o nome *pandéktes* (que abrange tudo) por conta da função de mediador entre o Nome e o Verbo ser identificado no particípio, deixando o advérbio como mediador de toda as partes do discurso.

A *Téchne Grammatiké* de Dionísio Trácio, um breve tratado metodológico de gramática, foi uma das primeiras obras sobre o ensino de língua e literatura, tal obra presente na edição de Bekker (1965 apud NEVES, 2005)² dá continuidade a tradição da escola de Aristarco e examina cada parte do discurso (frase) em separado no seu manual. O advérbio na sua obra é classificado como a parte não flexionada no discurso e se situa antes ou depois do verbo, mais explicitamente, Bekker diz que o advérbio se relaciona com a forma do verbo de uma maneira geral ou particular. Dionísio apresenta 26 classes do advérbio, diferenciando entre simples (*pálai*) e compostos (*própalai*) e, apresentados de forma denotativas. Neves (2005, p. 126 – 128) lista os seguintes itens pertencendo a essa classificação:

1. *Chrónou* (Tempo) – como *nýn*, *tóte*, *áuthis*. Remetem a tempo determinado, como "agora", "hoje" e "amanhã";
2. *Mesótetos* (Posição intermédia) – Equivalente a Modo, advérbios gregos em – *ós* (port. = mente);
3. *Poiótetos* (Qualidade) – *láx*, "com o pé", *pýx*, "com o punho", *botrydón*, "em cachos";
4. *Posótetos* (Quantidade) – *pollákis*, "muitas vezes";
5. *Arithmoû* (Próprios para indicar número) – *dís*, "duas vezes";
6. *Topikón* (Lugar) – *áno*, "de cima para baixo". Há três ("maneiras de ser", *schéseis*) possibilidades, lugar "onde", "lugar "para onde"e lugar "de onde", respectivamente: *oíkoí*, "em casa", *oikade*, "para casa"e *oíkothen*, "de casa";
7. *Euchês* (Desejo) – equivalente a expressão "oxalá"³;
8. *Schetliastikón* (Dor ou pranto) – *ioú* (interjeição de dor);
9. Arnéseos ou Apopháseos (negação) – *au*, "não", *oukhí*, *oudêta*, *oudamôs*;
10. *synkatathéseos* (Afirmção) – *nái*, *náikhi*;
11. *Apagoréuseos* (Interdição) – *mé*, "não";
12. *Parabolês* ou *Homoióseos* (Semelhança, aproximação ou assimilação) – *hos*, *hósper*, *eúte katháper*;

2. 2 Que faz referência à edição de Bekker de 1814.

3. 3 Algumas expressões ou palavras não foram traduzidas por Neves (2005), Neves (2011), algumas são expressas por aproximação de significado.

13. *Thaumastikón* (Admiração) – *babaî* (interjeição de admiração);
14. *Eikasmou* (Conjectura) – *isos, tákha, tykhón*;
15. *táxeos* (Ordem, disposição) – *heksês*, "em seguida";
16. *Athroíseos* (Coleção, ajuntamento) ex.: *árden*, "inteiramente", *háma*, "de uma só vez";
17. *Parakeleúseos* (Exortação ou instrução) – *eía, áge, phére* (interjeição de exortação);
18. *Synkríseos* (Comparação) – *mállon, hêtton* "mais, menos";
19. *Erotéseos* (Interrogação) – *póthen, penika, pôs*, "onde?, como?";
20. *Epitáseos* (Intensidade) – *lian, sphódra, pány, ágan, málista*;
21. *Syllépseos* (Reunião) – *háma*, "de uma só vez", *homoú*, "em conjunto";
22. *Apomotikón* (Negação por juramento) – partícula grega *má*, usada antes do nome da divindade por quem se jura;
23. *Katomotikón* (Afirmção por juramento) – partícula grega *né*, usada antes do nome da divindade por quem se jura;
24. *Bebaiéseos* (Certeza, consolidação) – *deladé*;
25. *Thetikón* (Imposição, obrigação) – *pleustéon*;
26. *Theiasmou* (Inspiração divina ou superstição) – *euoí*, "evoé".

Apesar de sua gramática possuir apenas os estudos fonéticos e morfológicos, a lista acima encontrada em Dionísio Trácio, e formalizada por Neves (2005), demonstra que o advérbio já recebia descrições que encontramos ainda hoje nas gramáticas ocidentais. Ressalta-se que as interjeições são nomeadas como um tipo de advérbio, mas foram posteriormente retiradas da classe dos advérbios.

Na obra *Da sintaxe das partes do discurso*, Apolônio Díscolo repete as classificações presentes em Dionísio o Trácio e da escola de Aristarco (NEVES, 2005), e afirma ordenar essas partes do discurso (frase) em uma proposição completa. Na ordem do discurso (frase), Apolônio coloca o advérbio como “uma espécie de adjetivo do verbo que deve vir após a proposição, que se liga ao nome” (NEVES, 2005, p. 173).

De acordo com Neves (2005), Apolônio demonstra que o advérbio é uma palavra invariável e argumenta que uma palavra variável empregada na posição de advérbio se torna indeclinável, a exemplo do *ir rápido*. Há a questão de alguns advérbios estarem juntos de qualquer verbo, outros em casos específicos. Apolônio não observa que o advérbio pode modificar estados e ações, mas define a natureza dos advérbios em primitivos e derivados, estes últimos derivados de advérbios, verbos, adjetivos, nomes etc. Apesar da tentativa de

abarcando os advérbios, Apolônio não tem êxito na regularidade da classificação, graças às características heterogêneas que não permitiam tal regularidade, algo que perduraria até hoje.

2.1.2 Nas gramáticas tradicionais

Antes de adentrarmos nas investigações linguísticas, é primordial observarmos a gramática tradicional devido ao seu caráter filológico e prescritivo perante as descrições da língua, mesmo possuindo objetivos completamente diferentes das investigações gerativas. Como visto anteriormente, o Advérbio é uma categoria usada para etiquetar itens de comportamentos diferentes, recebendo comparações como o adjetivo e a sua relação com o nome, que possui comportamento completamente diferente do advérbio. A nossa gramática, herdeira também da grega, com a mesma influência da Latina, perpetua essa problemática, a de classificar itens com comportamentos diversos em uma única classe, como veremos nas gramáticas tradicionais.

O léxico e a gramática de uma língua implicam uma organização categorial e quanto melhor as definições possíveis, melhor a clareza dos fenômenos da língua. Essa organização abarca diversos itens lexicais sob o rótulo “advérbio”, tornando a definição mais particular de autor para autor, mas mantendo alguns aspectos da vertente grega, definida por Dionísio Trácio, que afirma o advérbio como parte indeclinável, e nas indicações gramaticais sem flexões (de Caso, de tempo, de pessoa, de número e de voz), e de distribuição ao colocar o advérbio junto ao verbo (NEVES, 2005).

Isidoro de Serville (2006 *apud* PINTO, 2008, p. 290), um dos grandes estudiosos gramaticais na etimologia, conceitua o advérbio “pelo fato de que está acedente ao verbo, segundo se pensa, por exemplo, bene lege (ler bem). *Bene* é advérbio, *lege* verbo. Daí, então, é dito advérbio, porque sempre se completa junto ao verbo”. Com essa concepção, Serville identificava as marcas de significado do advérbio apenas em conjunto com o verbo, pois sem o verbo, o advérbio não possui sentido pleno. Um exemplo que ele aplica é na sentença *hodie scribo* (hoje escrevo), em que o *hodie* (hoje) possuirá sentido em conjunto do *scribo* (escrever). Os critérios que o autor aplica são mais intuitivos a respeito do advérbio e a sua relação com o verbo, e esse recurso empregado estará presente nos autores das gramáticas tradicionais do português como medida para destacar a heterogeneidade dessa classe.

Barros (1540), influenciado pela gramática grego-latina, mas possuidor de inovações e independência, buscou um caminho mais pedagógico e normativo sobre a moderna língua portuguesa no século XVI (MONTEIRO, 1997). O gramático caracteriza o advérbio como

acerca do verbo, demonstrando a necessidade de o advérbio estar em conjunto com o verbo para denotar eficácia ou remissão. O exemplo que ele mostra é com as sentenças (i) *eu amo a verdade*, (ii) *eu amo muito a verdade* e (iii) *eu amo pouco a verdade*. Para ele, o primeiro exemplo denota simplesmente *a obra de amar*; o segundo denota *a capacidade de amar* que o advérbio atribui a coisa amada e por fim; com o *pouco* se expressa *a diminuição do amar*, derrocando a capacidade de amar a coisa.

Na obra de Barros (1540), é concedido um poder de quantificar e qualificar o verbo, atribuindo uma característica de modificador ao advérbio. Os advérbios são classificados por espécie e figura, além da significação a partir das questões morfológicas. Por espécie, temos os primitivos (“muito” e “pouco”, por exemplo) e os derivados (“bom” deriva “bem”). Para os de figura, temos simples (ontem) e compostos (anteontem – significando antes de ontem). Já os de significação, Barros demonstra as inúmeras significações devido ao número de itens que se figuram como advérbio, por encontrar uma problemática em torno de não poder listar todos os possíveis, Barros mantém os conceitos gregos e latinos, reduzindo os advérbios em regras gerais de acordo com a ordem latina. Abaixo os itens listados nos respectivos tipos definidos por Barros (1540, p 60–61), vale ressaltar que algumas palavras estão traduzidas para o português moderno.

- Lugar – Aqui, ali, cá, lá, acolá, etc.,
- Tempo – Anteontem, ontem, hoje, agora, depois, cedo, etc.,
- Quantidade – Muito, pouco, mais, menos, maior, menor, etc.,
- Qualidade – Bem, Mal, etc.
- Afirmação – Certo, sim, etc.
- Negação – Não, nem, etc.
- Duvida – Quiça, por ventura, etc.
- Demonstrar – esse, este, esta.
- Chamar – ó, olá, etc.
- Desejar – ose, Oxalá;
- Ordenar – Depois, Idem;
- Pergunta – Como, sempre, etc.
- Juntar – juntamente, por em forma, etc.
- Apartar – À parte, afora, etc.
- Jurar – Certo, em verdade, etc.

- Despertar – Ei, etc.
- Comparar – assi, assi como, bem como, etc.
- Acabar – Em conclusão, finalmente, etc.

Além da lista dos advérbios que nela estão, Barros (1540) argumenta sobre os advérbios terminados em “-mente”, suprimidos da lista, afirmando que eles derivam de adjetivos femininos. Ele já adiantava uma das particularidades do advérbio na gramática tradicional, o advérbio ser derivado de adjetivos femininos. O autor não conceitua os advérbios como circunstâncias, como também, forma uma lista que incluíra as interjeições pertencendo aos advérbios.

No século XVI, a obra “Methodo gramatical para todas as línguas” (1619) de Roboredo (2007) demonstra técnicas da tradição greco-latina no ensino da língua latina para aprendizagem do português. Tal empreendimento se mostrou válido nos avanços de uma gramática mais próxima das ideias linguísticas e uma formalização melhor da gramática, além do método de auxílio aos alunos nas questões lógicas da língua.

Para Roboredo (2007), o advérbio é uma das cinco partes da oração (nome, preposição, verbo, advérbio e conjunção) e não possui número, regência e altera as palavras que se junta, em especial ao verbo, como também ao adjetivo (pertencente ao Nome) e a outro advérbio. O gramático divide sinteticamente os advérbios em duas categorias ou espécies, são eles, os advérbios de ação interna e os de ação externa. Seguindo Dionísio da Trácia, o gramático português inclui os itens de interjeição, mas define os mesmos por “significação acções interiores q os grãmaticos chamão, Interjeições, ou interposições, entre as palavras, como signaes naturaes do animo, não entrão em arte; porque os suspiros ou gemidos de cada hum servem em todas as linguas” (ROBOREDO, 2007, p. 71). Logo, ele mostra que esses itens fazem parte das ações interiores tão gerais que não podem ser particulares do vernáculo local.

Sobre a categoria externa, Roboredo (2007) classifica apenas as palavras em latim, demonstrando interesse em sistematizar os advérbios latinos e não do português, apesar das frases e suas traduções encontradas no final da obra. Ainda assim, não diminui a efetividade da classificação para a língua materna. Demonstramos apenas a classificação dos advérbios externos (Quantidade, Relação, Qualidade, Ação e Paixão e Lugar) e suas significações abaixo:

- Quantidade
 - Permanente
 - Tempo
 - Numeral

- Relação
 - Ordem
 - Repetição
 - Diversidade
 - Comparação

- Qualidade
 - Qualidade (modo)

 - Acrescentar
 - Diminuir
 - Apressar
 - Juntar
 - Dividir
 - Excluir

- Acção e Paixão
 - Perguntar
 - Responder
 - Afirmar
 - Confirmar
 - Negar
 - Mostrar
 - Duvidar

- Lugar
 - Onde
 - Donde
 - Para Onde

Na sua obra, ele trata explicitamente do advérbio no Português Europeu (PE) afirmando que poucos são os simples e os compostos são maioritariamente com final “- ente”. Os elementos lógico-formais do estudo de Roboredo (2007) mantém, apesar da criatividade na abordagem metodológica da sua gramática, uma vez que, a tradição acerca do advérbio marca o significado semântico e modificador dos diversos itens na sintaxe, porém, se mantém inalterado em relação as outras gramáticas que abordam esta categoria.

O advérbio nas gramáticas antes do século XIX conceituam o advérbio muito ligado as questões greco-latinas, como uma das partes do discurso e “palavra invariável”, com algumas diferenças de subcategorização. As definições vigentes separam em outra classe as interjeições e a classificação é muito menos incoerente entre si, apenas mantém a obscuridade devido às diferenças de conceituação entre autores, mas sem se afastarem de alguns elementos característicos dessa classe na Gramática Tradicional.

As gramáticas de linha mais filosóficas existentes no século XIX tinham como principal gramático Barbosa (1866). Para esse gramático, era preciso analisar a partir do uso o comportamento das palavras, para assim identificar, numerá-las e classificá-las. Por essas serem expressões do pensamento, era preciso organizá-las a partir das suas naturezas e funções. Continua Barbosa ao afirmar que as línguas servem, em primeira ordem, não à comunicação, mas ao raciocínio, pois cabe a elas organizarem o pensamento e a ordem dos seus constituintes para que cada parte do pensamento seja identificado e entendido. Tinha uma noção de gramática representativa dos pensamentos em forma de sons e palavras, assim como a literatura é representante da manifestação do ideal. Era a ciência dos *Reaes* (caracteres) representativos dos objetivos, ou dos *Nominaes*, o significado dos sons e das palavras. A gramática vista como ciência, deveria ser o estudo das representações mentais, pois a evolução da comunicação humana está inteiramente ligada a representação das ideias.

Um pensamento próximo dos postulados filosóficos da gramática de Port-Royal, o diferencia dos autores antecessores, por postular que as ideias vinculadas pela fala são organizadas na Língua. O gramático, nessa linha de raciocínio, deixa claro o lugar das interjeições, conceituando como linguagens primitivas que manifestam a emoção, transmitindo dor ou prazer, surpresas, etc. Sintaticamente postas como vocativos ou próximos a segunda pessoa da oração⁴. Tal distinção coloca as partes do discurso, segundo o autor, como objetos das ideias e sua ordem (sintaxe) à sua organização discursiva. Mesmo com essas inovações na gramática portuguesa, a sua gramática seguia um sistema que ele identificava como um sistema mais analógico de regras e exemplos do que lógico, com isso, acaba reconhecendo o ofício mais prescritivo do que analítico da sua obra.

Sobre o advérbio na gramática de Barbosa (1866), levando em consideração que as palavras para serem partes do discurso precisam ter funções próprias e independentes,

4. 4 A exemplo da sentença Ó pedro, vem cá, mas pode ser suprimida quando estiver no meio da sentença após o verbo: vem cá, (ó) Pedro. (BARBOSA, 1866, p. 103).

diferenciando uma das outras no discurso, afirma o autor “estes são huma Expressões Compostas, equivalentes a huma Preposição com seu complemento, que costuma ser hum Substantivo ou so, ou acom panhado de hum Adjectivo” (BARBOSA, 1866, p. 110), além de os advérbios terminados em *-mente* serem formados por adjetivos femininos. Essa concepção de atribuir o advérbio como uma preposição se encontra no próprio significado original do advérbio (*adverbium*, que significa “adjunto ao verbo”).

Por considerar a Preposição uma relação de complementos entre duas palavras, em que uma determina a outra, o advérbio estaria inserido neste contexto. Sendo uma redução ou expressão abreviada da preposição com seu complemento (uma palavra indeclinável), faz o advérbio agir igual, a juntar-se a qualquer palavra com mais concisão e brevidade, porém, não apenas ao verbo, mas a qualquer palavra suscetível de determinação, como os adjetivos e os próprios advérbios. Ao distinguir os tipos de advérbios, Barbosa (1866) os classifica em Advérbios propriamente ditos, Nomes Adverbializados e Fórmulas adverbias. Os advérbios propriamente ditos são os de redução da preposição com seu complemento em uma palavra só e invariável; Já os Nomes Adverbializados se diferem apenas por serem declináveis, porém suscetíveis em alguns contextos a forma indeclinável, como o adjetivo “certo” que varia para “certa” e receberiam número, mas o uso de “certo” no masculino, neutro e singular é empregado de forma invariável em “Certo sei” e “Certo que isto he malfeito”; As Fórmulas adverbias são as expressões suprimidas, quando separadas dos advérbios ou outro subentende a ocorrência do advérbio, como “d’aqui” e “d’ali” que significa, respectivamente, “deste lugar” e “daquele lugar”.

Com os Advérbios propriamente ditos do português, Barbosa (1866) concebe como hereditários do Latim e conceituam em advérbios de “lugar”, “tempo” e “quantidade”. Já os advérbios de “modo” e “qualidade” seriam vocábulos mais novos de formações românicas, como veremos na lista sintetizada por Saksida (2005, p. 82 – 83):

A – Formas hereditárias

- a) – Lugar: onde, algures, alhures, nenhures, aqui, ahi, dahi, aquém, além, cá, lá, acolá, arriba, cerca, dentro, fora, diante, traz (sic), longe, perto;
- b) – Tempo: quando, sempre, nunca, então, agora, avante, antes, depois, hontem, hoje, logo, já, ainda, cedo, azinha;
- c) – Quantidade: tão, quão, mui, mais, quase, cerca, apenas.

B – Formações românicas (portuguesas)

C – Advérbios de modo e qualidade: sim, não, assim, como, talvez, eis.

De fato, Barbosa (1866) trouxe algumas inovações sobre o advérbio, mesmo o seu modelo analítico passível de críticas, propôs soluções de alguns problemas e foi o primeiro a pontuar sobre a possibilidade de o advérbio não ser uma classe, mas uma subclasse, ao seu modo de ver, uma subclasse das Preposições por compartilharem algumas características, figurando de forma sintática o grupo dos Sintagmas Preposicionados. Além de reconfigurar os itens adverbiais em unidades maiores de modificação não apenas ao verbo, mas aos adjetivos, aos advérbios e conseqüentemente aos Nomes (substantivos). Como nos explica Saksida (2005) em relação às ideias de Barbosa em seguir coerentemente no que determina a palavra ser um fator elementar da oração, exercendo, no discurso, uma função exercida diferente das outras palavras, demonstrando não ser esse o caso do advérbio.

Com forte influência da tradição greco-latina e das teorias científicas darwinistas vigentes na época, Ribeiro (1885) afirma que a evolução das línguas estão ligadas as lutas por sobrevivência, pois são organismos sociológicos, decorrendo mudanças linguísticas mais perceptivas que as evoluções submetidas pelas espécies: “nenhuma língua parece ter vivido por mais de mil annos, ao passo que muitas especies parece terem-se perpetuado por milhares de seculos”. Ribeiro (1885, p. 136). Essa concepção influencia o gramático a conceituar o advérbio a partir dessa concepção linguística.

Para Ribeiro (1885), o advérbio é um dos oito grupos categoriais dos vocábulos da linguagem que expressa ou compõe as ideias do pensamento, com *status* de qualificadora – dependente de outra categoria – assim como o artigo e o adjetivo. O advérbio é uma palavra que qualifica um verbo, um adjetivo ou outro advérbio, e pertence ao grupo dos não flexionados, pois toma o traço +flexão e sua função para atribuir os tipos categoriais. Os artigos, substantivos, adjetivos, os pronomes e o verbo são variáveis; os advérbios, as preposições e as conjunções são invariáveis.

Podemos encontrar traços semânticos ao definir o advérbio como um modificador e o seu escopo delimitando a modificação perante as outras categorias já mencionadas. Dentre os subtipos de modificadores, Ribeiro (1885, p. 77) divide os advérbios de acordo com a sua natureza, em:

- Tempo – agora, ainda, amanhã, antes, cedo, hoje, antes, depois, já, jamais, logo, nunca, ora, quando, sempre, então, tarde;
- Lugar – onde, aqui, ahi, alli, aquém, além, acima, arriba, avante, cá, lá, acolá, afóra, dentro, algures, alhures, nenhures, longe, perto, longe, atrás;

- Ordem – primeiramente, ultimamente, depois;
- Modo – bem, mal, assim, como, acintemente, e a maior parte dos que se formam pela adjunção da terminação mente a um adjetivo;
- Conclusão lógica – conseguintemente, consequentemente;
- Quantidade – muito, pouco, assás, mais, menos, tão, quão, tanto, quanto, como, quasi;
- Afirmção – sim, verdadeiramente, efetivamente, realmente, certamente;
- Negação – nada, não, menos, nunca, jamais;
- Dúvida – talvez, acaso, quiça;
- Exclusão – só, somente, apenas, unicamente, siquer, sinão;
- Designação – eis.

O subtipo “Conclusão lógica” assume não apenas um modificador do verbo, de adjetivos ou de advérbios, mas pode ter outras funções como um conectivo entre sentenças ou um escopo de uma sentença inteira. Essa constatação poderia ser ampliada para todos os itens dessa lista, mas Ribeiro se limita apenas a classificá-las sem elaborar nenhum exemplo sobre o uso de tais advérbios.

O gramático postula que o advérbio se caracteriza como uma classe de entremeio por marcar a transição das palavras variáveis para as invariáveis. Tal argumento é sustentado por Ribeiro (1885, p. 135) ao afirmar que "admitir graus de comparação (*lindamente, mais lindamente, lindissimamente, boamente, melhormente, Optimamente*) revela o advérbio ter sido palavra flexional nas antigas linguas indo-germanicas, fontes - da portugueza", visto como uma parte não dependente em relação aos outros grupos categorias, sinalizando a complexidade de conceituar o advérbio.

Nas gramáticas do século XX, encontramos em Oiticica (1955), algumas inovações da época nas definições linguísticas logo no início de sua obra, considerando a linguagem como manifestação do pensamento e sentimento; a fala como um meio de veiculação da linguagem por meios sonoros e físicos; língua como sistema; a gramática como exposição dos fatos da língua e; a linguística como o estudo desses fatos. Dividindo os fatos gramaticais em dois, os de lexicologia e sintaxe, o advérbio é visto em níveis taxionômicos no seu escopo.

Seguindo as ideias greco-latinas, Oiticica (1955) define o advérbio como modificador do verbo, exprimindo circunstância que cercam ou definem uma ação. Para ele, os advérbios de intensidade podem prender-se a adjetivos e advérbios com função de índice de grau, a exemplo *muito rico = riquíssimo*. porém, o gramático não concebe as palavras *sim* e *não* e as partículas de negação como advérbios, mas sim, pertencendo a classe dos denotativos, classe

negligenciada pelos gramáticos, como *pois sim, não vê, nunca* etc. A sua lista de advérbios difere dos outros gramáticos, por não incluir justamente os subtipos negação e afirmação, como visto acima.

As palavras *apenas, sequer, senão, só, somente, exclusivamente, unicamente, menos, fora*, etc., aplicadas a frases: “apenas eu fiquei na rua”; “não vi um soldado na sequer”; “todos, senão meu pai, atravessarem o rio”; “só a moça não chorava”; etc, identificadas como advérbios de exclusão ou preposicionadas demonstram, segundo Oiticica (1955), a confusão e discordância dos gramáticos pela distribuição a esmo dos advérbios sobre diversos rótulos (subtipos). A exclusão (assim como a inclusão) só deve incidir sobre o substantivo, e não fazer referência ao verbo, logo, não pode ser caracterizado como circunstancial, portanto, não pode ser advérbio.

Essa incoerência de classificar palavras que incidem sobre o substantivo serem classificadas como “advérbios”, ao ver do gramático, há circunstancialidade quando afeta o verbo, é preciso rever tal taxionomia para a palavra classificada como advérbio diante de substantivos. Proposta essa que Oiticica (1955) coloca em prática perante a insuficiência taxionômica, na tentativa de classificar o inclassificável, na gramática tradicional. Assim ele observa que essa problemática torna trabalhoso o ofício do professor ao se deparar com essa falha dos antigos gramáticos. Para ele, a urgência de proceder cientificamente para solucionar tal problema deve seguir os fatos da linguagem, distinguindo três classes de palavras que exprimem movimentos e operações mentais, são elas “ideativas”, emotivas e denotativas. As palavras nominativas, pronominativas, modificativas e conectivas devem fazer parte das ideativas, além das palavras sintéticas; as interjectivas e intensivas revelam emoções puras, emoções que reforçam uma ideia ou ênfase de uma palavra, logo fazem parte das emotivas; as denotativas subdivide-se em aditivas, afirmativas, avaliativas, comparativas, concessivas, corretivas, designativas, distributivas, escusativas, exclusivas, expletivas, explicativas, gradativas, inclusivas, preventivas, seletivas.

Reorganizando as possibilidades de classificação elaboradas por Oiticica (1955), o advérbio recebe traços semânticos (modificativa) e de escopo (componente sintático – junto ao verbo e com certas restrições nos adjetivos ou em outro advérbio), mantendo a característica principal de ser invariável. Percebemos a distribuição de advérbios pelas classes de palavras (ideativas, emotivas e conotativas), principalmente nas conotativas. Essa distribuição reduziu o número dos advérbios gerais em oito subtipos, seguindo os itens do grupo modificativo (semântico): dúvida, frequência, intenção, intensidade, lugar, modo,

ordem e tempo. Reproduzindo tal classificação semântica abaixo apenas dos advérbios limitados às palavras ideativas, temos em (OITICICA, 1955, p. 56):

Modificativas:

- Dúvida – talvez, quiçá, acaso, por ventura;
- Frequência: diariamente, quotidianamente, semanalmente, mensalmente, nunca, jamais, sempre, às vezes, raramente, uma vez, sucessivamente, constantemente;
- Intenção – acinte, acintemente, adrede, intencionalmente, propositadamente, premeditadamente, acaso, casualmente;
- Intensidade – muito, assaz, bastante, excessivamente, demais, demasiadamente, pouco, mais, menos, tão, tanto, quão, quási, meio, depressa, devagar, demoradamente.;
- Lugar – abaixo, acima, arriba, aquém, além, aqui, ali, cá, lá, acolá, avante, atrás, algures, alhures, nenhures, diante, detrás, dentro fora, longe perto, onde;
- Modo – atoa, bem, certo, mal, errado, tristemente (incluí diversos adjetivos adverbializados com o sufixo –mente);
- Ordem – primeiramente, anteriormente, depois, posteriormente, antes, atrás, adiante;
- Tempo – ainda, agora, amanhã, dantes, cedo, então, hoje, ontem, já, logo, tarde, outrora, imediatamente, etc.

Apesar de reorganizar os advérbios de acordo com as características semânticas e apenas um escopo, Oiticica (1955) distribui, o que seriam advérbios em outras gramáticas nas denotativas por ver a impossibilidade de modificarem o adjetivo, o advérbio e o verbo por não serem circunstanciais, a exemplo, os de negação e afirmação, identificadas como denotativas afirmativas ou negativas.

Para Said Ali (1965) o advérbio denota um evento circunstancial de tempo, lugar, modo, grau ou intensidade, negação, dúvida, etc., determinando o verbo, o adjetivo e outro advérbio. Expresso por meio de uma palavra invariável ou locução de sentido equivalente. Ainda para o gramático, os advérbios interrogativos são representados por palavras invariáveis que assumem o sentido de tempo, lugar, modo ou causa quando inseridas em perguntas. Também recebem destaque os advérbios terminados em "-mente", derivados de adjetivos femininos, e os advérbios suscetíveis a gradação, assim como os adjetivos comparativos e superlativos.

A síntese da classificação do advérbio de acordo com seu traço semântico postulado por Said Ali (1965, p. 89 - 91) pode ser vista abaixo:

1. Circunstanciais:

- a) Lugar: êle trabalha *aqui*, em S. Paulo;
- b) Tempo: êle trabalha *hoje*, todos os dias;
- c) Modo: êle trabalha *aplicadamente*, com aplicação;
- d) Negação: êle *não* trabalha;
- e) Dúvida: êle *talvez* trabalha;
- f) Grau ou intensidade: vejo-te *tão* alegre; fiquei *muito* triste; chegaste *mais* tarde.

2. Interrogativos:

- a) *Quando* chegaste a Capital?
- b) Não me disseste *quando* chegaste.
- c) *Onde* achou êle a bengala?
- d) De *onde* vens tu tão cansado?
- e) *por que* não respondes à minha pergunta?
- f) *Como* se pronuncia esta palavra?
- g) Queria saber *como* se pronuncia esta palavra.

3. Advérbios em -mente⁵:

- a) Êle portou-se *corajosamente*;
- b) Entendemos *fácilmente* o que lemos;
- c) Os revoltosos penetraram na sala *ameaçadoramente*.

Exceção dos adjetivos terminados em -ês: burguês – burguêsmente.

4. Grau ou Intensidade:

Principais – tão e tanto, quão e quanto, quase, muito, pouco, mais, menos, bastante. Tão e Quão como determinantes, como adjetivos de forma positiva, como quantitativo ou como outros advérbios e Tanto e Quanto combinando com o grau comparativo:

5. Said Ali (1965, p. 89) afirma que esses advérbios são de modo e derivam dos adjetivos e dos adjetivos femininos quando há flexão de gênero no adjetivo.

- a) *Tão* lindos olhos;
- b) Vestidos *tão* ricos;
- c) Marília tem *tão* poucas jóias;
- d) *Quão* preciosos;
- e) *Quão* poucos;
- f) Chegaste *tão* cedo;
- g) Leitura *tanto* mais difícil;
- h) Dias *tanto* piores;
- i) Ação *tanto* menos nobre;
- j) *tanto* maior;
- k) *Quanto* mais longas.

5. Gradação dos advérbios:

Comparativo e superlativo:

- a) Cheguei *tão depressa* como o correio;
- b) Andamos *menos devagar* do que eles;
- c) Ligado *estreitíssimamente* ou *muito estreitamente* a sus deveres.

Familiar -inho, -zinho:

- a) Levanta-te cedinho (= muito cedo);
- b) O doente está melhorzinho (= um tanto melhor).

Tal distribuição baseada em uma concepção semântica, tornando o advérbio um circunstancial, o coloca como um modificador (acessório ou particular) do verbo, do adjetivo ou de outro advérbio. Porém, não encontramos maiores explicações para essa ocorrência além da listagem seguindo as circunstâncias envolvendo o escopo. Em Said Ali (1971), é marcada a possível origem de alguns detalhes deixados de lado, como o fato do advérbio acrescentar conceitos de tempo, lugar, modo etc., delimitando-os ou esclarecendo o sentido sem exercer função de complemento. Além de afirmar que a maioria dos advérbios portugueses não provêm exatamente do latim, mas sim, das formações desconhecidas do latim literário, com novas criações, especialmente a criação de advérbios em *-mente*. Essa terminação surge do ablativo latino *-mens* que passa a ser *-mente* na nossa língua. As locuções adverbiais são combinações de preposições com substantivos exercendo a função de advérbios, só que ao contrário de *-mente* que reduz por economia das frases, as locuções são mais longas.

Said Ali (1965) afirma que o Adjunto adverbial, termo acessório formado por um advérbio ou locução adverbial, acrescenta ao predicado o esclarecimento de lugar, tempo, modo, etc. Não se detendo, além disso, por ver a posição do advérbio já marcada de acordo com a sua morfologia e semântica.

Esse mesmo raciocínio é expresso por Câmara Jr (1999, p. 79) ao afirmar que em “face dessa divisão morfo-semântica, a classificação funcional subdivide nomes e pronomes pela sua função na comunicação lingüística, [...] de natureza funcional também, é o advérbio. Trata-se de um nome, ou pronome, que serve de determinante a um verbo: *fala eloqüentemente, fala aqui*”. Conceituando, o advérbio é termo determinante apenas do verbo, com função complementar, na língua, acrescenta uma qualificação ao adjetivo, por exemplo: muito feliz, efusivamente feliz. Porém, essa função não é geral a todos e não devem entrar na definição essencial da função do advérbio.

Como foi observado, os gramáticos tentaram estabelecer critérios que definiam os advérbios como uma classe sem flexão, invariável e na sua maioria, derivados dos adjetivos femininos ou adjetivos masculinos não flexionados. Apesar de terem conceitos próximos em relação aos advérbios, a sua heterogeneidade semântica fez que classificassem os itens de acordo com o comportamento que exerciam (tempo, modo, dúvida, circunstância, etc.) na visão desses gramáticos, mas sempre pautados na gramatical greco-latina que definiria o advérbio no que ele é hoje na GT. Com análises mais detalhadas e reflexivas, apresentamos Bechara (2009) e a perspectiva da gramática moderna sobre o advérbio para tentar esclarecer um pouco mais esta categoria nos estudos da língua portuguesa.

Influenciado por Said Ali e pela linha estruturalista de Mattoso Câmara Jr, encontramos em Bechara (2009) inovações advindas dos estudos linguísticos do Séc. XX e início do Séc. XXI, a exemplo dos estudos funcionalistas, porém, mantém a linha clássica dos grandes gramáticos em abordar a norma culta em suas páginas. Repetindo o conceito sobre o advérbio de Said Ali, o advérbio é a expressão modificadora que denota uma circunstância de lugar, de tempo, modo, intensidade, condição, etc., e desempenha sintaticamente a função de adjunto adverbial, além de ser constituído por palavra de natureza nominal ou pronominal e se refere ao verbo, e dentro de um grupo nominal unitário, a um adjetivo e a um advérbio, ou a uma frase inteira (BECHARA, 2009).

Por ser uma classe heterogênea, não é possível uma classificação uniforme e coerente, o gramático observa que há uma flexibilidade de posição devido ao advérbio não apenas se prender ao verbo e a extensão do predicado, mas anteceder ou ficar posposto ao sujeito. Com essas possibilidades, há advérbios semânticos-sintáticos mais internos ligados ao núcleo

verbal com menos flexibilidade de posições e os mais externamente ligados ao núcleo verbal, dificultando qualquer projeto de classificação unitária para todos os casos. Por isso, a melhor maneira indicada por Bechara (2009) é ater-se a cada relação circunstancial dentro do enunciado/frase em suas diversas possibilidades. Com isso, os valores semânticos são identificados e classificados nos advérbios de tempo (agora, antes, tarde, depois, etc.), de lugar (aqui, fora, ali, etc.), de quantidade (tanto, muito, bastante, etc.), dentre muitos outros. Por outro critério, teremos os advérbios funcionais demonstrativos (aqui, então, agora, aí, etc.), relativos (onde, como, quando, etc.) e interrogativos (quando?, onde?, como?).

Além da classificação, Bechara (2009, p. 245) lista as principais circunstâncias expressas pelo advérbio e pelas locuções adverbiais:

- Assunto: Conversar sobre música.
- Causa: Morrer de fome.
- Companhia: Sair com os amigos.
- Concessão: Voltaram apesar do escuro.
- Condição: Só entrará com autorização. Não sairá sem licença.
- Conformidade: Fez a casa conforme a planta.
- Dúvida: Talvez melhore o tempo. Acaso encontrou o livro.
- Fim: Preparou-se para o baile.
- Instrumento: Escrever com lápis.
- Intensidade: Andou mais depressa.
- Lugar: Estuda aqui. Foi lá. Passou pela cidade. Veio dali.
- Modo: Falou assim. Anda mal. Saiu às pressas.
- Referência: “O que nos sobra *em glória* de ousados e venturosos navegantes, mínguanos *em fama* de enérgicos e previdentes colonizadores”.
- Tempo: Visitaram-nos hoje. Então não havia recursos. Sempre nos cumprimentaram. Jamais mentiu.
- Negação: Não lerá sem óculos.

O conceito dessa classificação subentende um significado discursivo circunstancial, em que o advérbio pode ser inserido nas expressões pontuais que lhe configura o tipo expressado, ou seja, o tipo se forma a partir da sua função circunstancial. Uma outra característica evidenciada por Bechara (2009) é o fato de certos advérbios agirem como unidade gramatical

de retomada ou substituição de um discurso presente ou virtualmente presente, ou previsto no diálogo. Já é previsto na gramática tradicional que o pronome exerce esse papel, porém, graças a ação da antitaxe que o papel textual dos advérbios *sim, não, talvez e também* deixam de ter função de advérbio e adquirem a função de substitutos quando retomam respostas textuais rápidas, como nos mostra Bechara (2009, p. 256) logo abaixo:

Você vai ao cinema? – Sim.

Ela fez os exercícios? – Não.

Tu não foste escolhido? Também.

Já os advérbios em *-mente*, segundo o gramático, ao saírem da camada no nível da palavra para funcionarem no nível da cláusula e, conseqüentemente, no nível da oração ou texto, exercem função na camada superior, ou seja, saem da camada inferior para a superior exercer uma oração ou texto, como veremos a baixo os exemplos do autor:

Certamente!

Naturalmente!

Além dos de nível da oração e do texto, também os que estão no nível da cláusula comentário:

- *Certamente* ela não virá hoje.
- Todos saíram ilesos, *felizmente*.
- *Naturalmente* ele negará o que disse ontem.

Esses exemplos demonstram que a característica canônica do advérbio ser exclusivamente preso às referências do núcleo verbal não se aplica. Indo além das características canônicas do advérbio, Bechara (2009) identifica o papel delas em diferenciar orações com valor existencial, atribuído ao estado de coisas indicado pela oração, como os advérbios de negação e dúvida ao determinar existência certa, negada, duvidosa, desejada, dentre outros.

- Ele veio / Ele não veio
- Ela chega / Ela talvez chegue

Os advérbios que exercem determinação a nomes e pronomes são identificados como nominais e pronominais. A observação feita por Bechara (2009) é que os nominais são formados a partir de adjetivos acrescidos de *-mente*, se posicionando mais ou menos entre a derivação e a composição de um advérbio terminado em *-mente*, fonológica e morfologicamente falando. Por exemplo: rapidamente (= de modo rápido). Alguns pronominais são:

- Demonstrativos: aqui, aí, acolá, lá, cá.
- Relativos: onde (em que), quando (em que), como (por que).
- Indefinidos: algures, alhures, muito, pouco, que.
- Interrogativos: onde?, quando?, como?, por que? (por quê?)

Vimos em Bechara (2009) alguns conceitos recentes referentes ao advérbio, principalmente em reconhecer e demonstrar a possibilidade de o advérbio funcionar como Nome e Pronome, modificando um substantivo. Apesar de trazer dúvidas quanto a classificação de algumas palavras ao exercerem circunstância, como os de modo e, em outro contexto, exercer a função de adjetivo ao se referir a um substantivo, não há tentativa de formalização dos advérbios, nem a preocupação de demonstrar o funcionamento do escopo, tendo em vista que ele parte de um viés funcionalista. Ao falar sobre a locução, o gramático define a locução adverbial como duas ou mais palavras que funcionam como advérbio, um conceito clássico já visto por outros gramáticos anteriores. Não é claro, no entanto, se tais expressões são fixas para que sejam classificadas como advérbio, problema esse ainda perpetuado pela gramática tradicional.

A falta de formalização e a classificação ser inteiramente pautada no fator circunstancial demonstra a dificuldade em delimitar a classe gramatical do ponto de vista lexical, sem definir o que distingue advérbios de locuções, por exemplo. Outra questão é em relação ao grau atribuído pelo advérbio, que o coloca sobre a mesma característica do adjetivo, mas posto na vaga noção de circunstância, somada a relação intensificadora gradual comparativa ou superlativa de acordo com as regras dos adjetivos. Vale salientar que Bechara (2009) não incluiu o advérbio de afirmação nos subtipos do advérbio.

Em síntese, o advérbio é classificado essencialmente como modificador do verbo e, em alguns casos como intensificador do adjetivo e de outro advérbio. Em alguns casos modifica o substantivo, se tornando um modificador nominal e pronominal. Sua tipologia é referente ao seu valor semântico, atribuindo ao seu escopo o fato de ser circunstancial e

modificador do verbo; se estender com valor textual e gramatical, na oração ou texto, em substituição ou antecipação por clausura.

Observando as classificações e conceitos dos advérbios até aqui entre os gramáticos, é evidente que os advérbios estão divididos em dois tipos, os consensuais e os não consensuais. Dentre os consensuais estão os advérbios de Lugar, Tempo, Modo e Dúvida. Os advérbios não consensuais, são relativos as relações existentes com o adjetivo e com o próprio advérbio, tanto que alguns gramáticos admitem o advérbio como parte dos modificadores (preposicionados ou determinantes) do substantivo. Não precisamos nos repetir sobre a dificuldade que os gramáticos encontram diante de algo tão heterodoxo como os advérbios, pois nessas concepções tradicionais há diversas concepções encontradas entre gramáticas e autores, até autores não se satisfazem com apenas uma definição por tão vasto e diferenciados são os elementos presentes.

De modo geral, os valores semânticos e discursivos predominam diante de uma análise sintática ou de descrição teoricamente sistematizada das propriedades dos advérbios. Como indica Oliveira (1996), o maior problema visto até aqui, recai sobre os conceitos tradicionais do Advérbio que não servem propriamente à Linguística, já que os vários tipos de itens denominados Advérbios formam uma só Categoria. Se olharmos para a Gramática Gerativa (GG), por exemplo, a GT se preocupa com uma descrição elaborada, já que se preocupa em descrever uma língua padrão idealizada e prescritiva, além do que discutimos aqui, as questões da metalinguagem explorada até o momento, por esse motivo, buscamos uma proposta mais descritiva antes de observamos as discussões de cunho gerativo. Por isso, a próxima seção conduz o advérbio na perspectiva descritiva.

2.2 QUESTÕES DESCRITIVAS

Nosso último gramático a trazer algumas considerações a respeito dos advérbios é Perini (2005)⁶, pelo seu empenho em expor os estudos gramaticais a partir das pesquisas advindas da linguística e propor a explicitar os problemas e oferecer soluções para a maioria dos casos nela existentes, assim afirma que o

6. Deixamos de lado os estudos críticos da Bomfim (1988), além de outros gramáticos contemporâneos como Ataliba Teixeira de Castilho, com a sua Nova Gramática da Língua Portuguesa por questão de espaço e tempo.

mais importante é suscitar discussão — discussão de problemas reais da análise da língua, baseada em fatos observáveis do português padrão atual. Meu objetivo é contribuir para uma reorientação radical dos estudos de língua portuguesa: do exame das gramáticas para o exame dos fatos da língua - sem desprezar o exame das gramáticas, mas submetendo-as a um crítico rigoroso (2005, p. 15).

Em geral, esse objetivo passa por alguns problemas em relação ao número de fenômenos da língua que a gramática cobre em comparação a linguística. Isso é evidente ao descrever os itens classificados como advérbios, em que as descrições circulam por critérios semânticos e sintáticos com certo nível de irregularidades, porém, sem os detalhes mais profundos advindos da gramática descritiva, a conclusão sobre esta classe se torna duvidosa. Diante desses impasses, Perini (2005) propõe uma maneira nova de descrever a estrutura do português a partir de princípios teóricos mais rígidos, apontando, também, os acertos e equívocos da GT.

Existe uma classe dos advérbios? Questiona Perini (2005) ao observar que a categoria tradicional dos advérbios possui uma série de classes que se diferenciam muito no comportamento sintático, embora a prioridade da GT é definir tal classe a partir dos traços semânticos (modificar outros itens e a si mesmo), ocasionando um elemento circular na sua definição clássica. Já que se define uma classe pela função que exerce, temos uma classe multifuncional para classificar como tal. O gramático ilustra as potencialidades funcionais dos advérbios separando em alguns grupos, evidenciando não só as inconsistências semânticas, mas também sintáticas, os seguintes itens classificados como advérbios *não*, *rapidamente*, *completamente*, *muito*, *francamente*⁷ e subclassificados semanticamente como *de negação*, *de modo* e *de intensidade*, ao analisar tais palavras o autor nos mostra que as funções sintáticas também são diversas. Como veremos, essas palavras, de acordo com o gramático, ocupam pelo menos seis funções, e alguns podem ocupar mais funções como podem ver nos traços funcionais⁸ elencados:

1. Negação verbal:

a) Seu tio [+NV]*não* apareceu na estação.

7. 7 Perini (2005, p. 87) afirma que na “literatura linguística moderna, elementos como francamente são considerados à parte, como um elemento anexo à oração, talvez mesmo externo a ela; denominam-se ‘advérbios de oração’”.

8. 8 O significado das abreviações funcionais: AA = adjunto adverbial; AC = adjunto circunstancial; Ant = anteposição; AO = adjunto oracional; Atr = atributo; Int = intensificador.

2. Intensificador:

- a) Almeida é [+Int+AC] *muito* magro;
- b) Almeida estava [+Int+AA] *completamente* bêbado;
- c) Essa proposta é [+Int,+AO,+Atr] *francamente* ilegal.

3. Adjunto circunstancial:

- a) Ela ri [+Int+AC] *muito*.

4. Atributo:

- a) Terminamos a pintura [+Atr] *rapidamente*.
- b) Ela me revelou tudo [+Int,+AO,+Atr] *francamente*.

5. Adjunto adverbial:

- a) Ela decorou o apartamento [+Int+AA] *completamente*.

6. Adjunto oracional:

- a) Acho que ele nos enganou, [+Int+AO+Atr] *francamente*.

Seguindo a observação de Perini (2005), a classificação tradicional não analisa essas diferenças encontradas nos itens em relação ao comportamento gramatical. Tampouco observa que os itens pertencentes às subclassificações (de modo, de intensidade etc.) se mostram incoerentes as classificações sintáticas vistas acima, exemplo dos advérbios de modo (*rapidamente*, *completamente* e *francamente*) que, segundo o autor, são sintaticamente diferentes. Além de alguns compartilharem traços encontrados em outras classes (o traço [+AC] encontrado nos adjetivos que exercem a função de adjuntos circunstanciais) demonstrando que não é uma classe com todos os itens exclusivos.

Sobre a noção de modificação, apesar de vaga, Perini tenta solucionar tal impasse afirmando que a

noção de “modificação” é bastante obscura; como a interpreto, seria um misto de semântica e sintaxe. Semanticamente, modificação “significa que um advérbio teria seu significado amalgamado ao de um outro elemento, formando um todo semanticamente integrado; assim, digamos, *corremos* exprime uma ação, e *corremos depressa* exprime a mesma ação, crescida de algum ingrediente de significado. Tanto *corremos* quanto *corremos depressa* seriam unidades no plano semântico. (PERINI, 2005, p. 340) [grifos do autor]

Apesar de reconhecer que é vaga tal hipótese e poder aplicar com qualquer outra classe essa definição, aparentemente, a noção de modificação na sintaxe está relacionada ao conjunto de constituintes que se relacionam entre si. Assim como *corremos* se relaciona com *depressa*, mas, ainda assim, é vaga, pois a outros casos que Perini aponta que ocorrem com outras classes (*Comi uma peixada* – onde *comi* e *uma peixada* estão em construção, mas *uma peixada* não é uma constituinte adverbial). Essa determinação se torna difícil, já que definindo o advérbio como elemento modificador do verbo qualquer construção seria adverbial.

Não é garantia que o “estar em construção” seja uma característica das funções sintáticas, como afirma Perini, também não há garantias de que um item lexical seja modificador, pois o que se julga é uma lista de palavras assim chamadas. Observamos o exemplo:

(1) *Somente* André percebeu a situação.

Verifica Perini que a palavra *somente*, considerada advérbio, modifica o substantivo, André. Por outro lado, a sentença:

(2) Ela escreve *rápido*.

O adjetivo *rápido* que está modificando um verbo, colocando uma pá de cal nas pretensivas definições exatas sobre os elementos que ele modifica. A conclusão que Perini chega é que os elementos em construção com o verbo ou adjetivo etc. não são critérios para definir alguma classe, no possível, deve ser definido em critérios de função. Diante disso, a classe dos advérbios é uma lista de classes ou subclasses e essa lista ainda não foi até o momento, segundo o autor, sintetizada para termos um panorama maior.

Esse problema se estende ao Sintagma Adverbial, segundo Perini (2005), a simplicidade do conceito e as funções múltiplas correspondentes no AdvP seriam de atributo, adjunto adverbial, adjunto oracional, negação verbal, adjunto circunstancial. Cada uma delas se mostrando incoerentes na uniformidade de funções, tendo em vista a negação verbal, que desempenha outras funções não relacionadas como os outros constituintes do AdvP. Vejamos os exemplos (ídem, p. 120):

(3) a. *Com franqueza*, acho que eles nos enganaram.

b. Ele se expressou *com franqueza*.

Em (3a), o sintagma *com franqueza* desempenha função de adjunto oracional, enquanto (3b) desempenha função de adjunto circunstancial. A crítica de Perini aos conceitos gramaticais tradicionais se repete pela falta de uma abrangência suficiente para definir (não só) o Sintagma Adverbial (AdvP, do inglês *Adverb phrase*), já que até o momento, o que se tem é um conjunto de funções presentes no AdvP que podem ser formados até por preposição e

Sintagma Nominal (NP, do inglês *Nominal Phrase*) desempenhando função adverbial, por exemplo.

Por fim, Perini ao falar sobre a GT no seu percurso em relação aos advérbios, afirma que a GT não encontrara consonância a respeito das subcategorias e tipos nessa categoria, tampouco realizaram uma exaustiva bateria de testes para dar conta de tal heterogeneidade. Se atendo ao seu objetivo prescritivo das gramáticas normativas, em dois momentos percebemos o tipo de abordagem que figuraram tais estudos, primeiro, em uma época mais antiga se classificava por critérios mais intuitivos, no século XX em diante, por critérios mais explícitos, de cunho mais funcionais e propriedades semânticas, principalmente sobre as evidências em torno dos fenômenos ligados ao advérbio agindo como modificador do verbo. Hoje prevalece a ideia de emitir circunstância, ser invariável e modificar o constituinte referente mais próximo.

2.3 O ADVÉRBIO NOS ESTUDOS DA GRAMÁTICA GERATIVA

Ao ocupar-se dos Advérbios, o Gerativismo encontrou algumas divergências em focalizar tanto a diversidade e a parametrização, quanto descobrir as regras gerais que caracterizam, de modo integrado, o comportamento do advérbio. Algumas tentativas de reclassificação ou de redução do quadro herdado pela tradição gramatical por vias mais sintáticas, compreendia uma gama de classificações que foram feitas, segundo Oliveira (1996), entre os Advérbios disjuntos ou de frase (4a) e; Advérbios adjuntos ou de constituintes (4b):

- (4) a) *Basicamente* eu não posso interferir no processo global... mas eu queria entender esse processo.
 b) O Brasil diz-se *basicamente* subdesenvolvido e diz-se também que ele está crescendo.

Com um viés mais semântico, são classificados como Modo (modificando o verbo) ou como predicador de predicadores; orientados para o sujeito, funcionando como predicadores de dois argumentos; e orientados para o falante. Além de outra classificação que agruparia os Advérbios em focalizadores, modalizadores e aspectuais (ibidem, p. 25).

- Focalizadores:
 - É *justamente* disso que não precisamos.
- Modalizadores:
 - Precisa *realmente* estar convencido.
- Aspectuais:
 - *De vez em quando* ele chega a necessitar de agrônomos

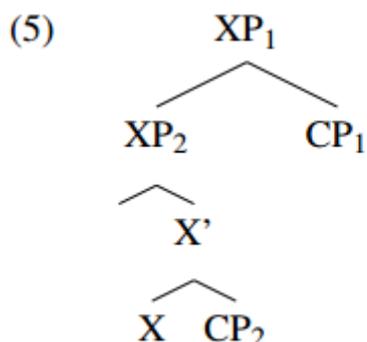
Oliveira (2010) afirma que há evidências em diversos trabalhos de autores que partem do P&P (Princípios e Parâmetros) defendendo que não existe “uma definição de traços específicos para a codificação de um sintagma adverbial (AdvP) no léxico na mente humana e que os sintagmas adverbiais são simples casos da categoria adjetivo” (ibidem, p. 210). Tal ideia que já era introduzida pelo gramático português Barbosa (1866), não pode ser descartado, apesar de algumas evidências de não complementação semântica entre as duas classes, além de ter diferentes comportamentos sintáticos em muitos casos, a coordenação entre eles não se faz possível em outros casos, ou seja, alguns advérbios não podem modificar o adjetivo. Como veremos mais a frente a partir de Jackendoff (1972) em diante.

Diante dessa rápida apresentação de alguns questionamentos, mostraremos a trajetória dos estudos sobre o advérbio, iniciando por pesquisas, em linhas gerais, do advérbio como adjuntos, seguindo posteriormente as pesquisas de Jackendoff (1972) até a proposta dos Sintagmas Adverbiais como Especificadores, em Cinque (1999), além de alguns estudos mais recentes sobre o tema. Com o propósito de mensurar os estudos sobre o advérbio, nossa intenção nessas seções que seguem é simplesmente a observação de alguns dos diferentes estudos que buscam dar conta dos advérbios. Com isso, as pretensões de reformulação sobre elas inexistem, pois as evidências aqui servirão como possibilidades de evidenciar e formalizar melhor os advérbios nos estudos de cunho gerativo que alimentam os trabalhos ligados ao processamento natural da linguagem (PNL), visto que o mesmo carece de trabalhos mais específicos que tratem dessa classe.

2.3.1 Advérbios como Adjuntos

Os adjuntos, segundo Míoto, Silva e Lopes (2013), são os tipos de sintagmas licenciados sem ser considerados complementos ou especificadores de núcleo. Este princípio de projeção adjunta é formada por duas noções adicionais, a Inclusão (α inclui β se e somente

se todos os seguimentos de α dominam β) e a Continência (α contém β se nem todos os segmentos de α dominam β). Continuando, Miotto, Silva e Lopes (2013) representam com um esquema arbóreo em (5):



Essa estrutura, mostra o CP1 incluído no XP por ser dominado pelos segmentos XP1 e XP2; partindo do XP1 ou do XP2 ao CP2 em sentido descendente. Sempre o adjunto é representado pela duplicação de um XP ao qual ele será adjungido, já que ele está contido em uma única categoria.

Entre os pesquisadores que concebem a adjunção um fator primordial para os advérbios está João Costa (1996, 2004). De acordo com Gonzaga (1997), Costa (1996) tem por objetivo apresentar argumentos contra a adjunção à direita. Distingue o advérbio em dois tipos de interpretação (modo e orientado ao sujeito) e analisa-os de forma a encontrar evidências a sua tese, afirmando que a opcionalidade de adjunção à esquerda e à direita é baseado na ausência de alterações semânticas. Um dos problemas evidenciados por Costa (1996) está na impossibilidade de vários advérbios ocorrerem em posição final e em adjunções simultâneas, como no exemplo:

- (6) *John has spoken to his mother nicely carefully cleverly⁹.
- a. John spoke carefully to his mother (*nicely);
 - b. John spoke nicely to his mother (*carefully).

Diante do problema apresentado por Costa (1996), o autor se coloca a responder por que a interpretação dos advérbios (em alguns casos ambíguos) não são sempre variáveis, apesar de os significados poderem sofrer interferência de certos processos sintáticos. Costa

9. *João falou para sua mãe boamente cuidadosamente inteligentemente. Trad: Autor.

interpreta a adjunção do AdvP como uma forma de colocar o IP como complemento desse sintagma, além de não encontrar evidências que demonstrem que há a derivação dos advérbios que modificam outros advérbios neste tipo de estrutura, refutando assim a hipótese do movimento do sujeito para Spec de AdvP e rejeitando a interpretação dos advérbios orientados para o sujeito. Sendo assim, para ele, a única análise que posiciona os advérbios em posições pré-verbais são os que considera AdvP como adjunto de IP à direita e à esquerda.

A ideia de as projeções dos Adv adjungirem em diversas possibilidades, Costa (2004) explica que essas possibilidades têm a ver com a movimentação do Verbo para as projeções funcionais, como observados nas ordens do PE (COSTA, 2004, p. 7):

[AgrsP O João [TP leu [VP tV o livro]]]

Com a introdução do advérbio de tempo “ontem”, verifica, segundo o autor, se ele é adjungido em TP, assim como no inglês: S-Adv-V-O.

[AgrsP O João [TP ontem [TP leu [VP tV o livro]]]]

Também é adjungido como no francês, para o TP, mas na ordem S-V-Adv-O:

[AgrsP O João [TP leu [VP ontem [VP tV o livro]]]]

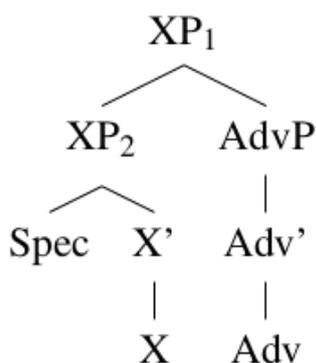
Esses exemplos são plausíveis para Costa, pois a sintaxe visível não satisfaz as condições para as interfaces, como nos lembra Santos (2011), pois a própria sintaxe visível gera uma saída com arranjos bem formados e avaliadas pelas interfaces, caso não sejam bem formadas, são bloqueadas na própria interface. A sua tese vai de encontro à perspectiva de Laenzlinger (1998), por argumentar a favor da redução da sintaxe visível (também discorda de Cinque (1999) veementemente), afirmando que o sujeito em PE move-se para preencher a posição de Spec nas três posições possíveis [AgrP [TP [VP]]] e o verbo apenas checa seus traços até o núcleo T^0 , não sendo possível se elevar a posições de núcleos funcionais acima, como os núcleos funcionaisAgr0. Assim, os advérbios, como afirma Costa (2004), são adjungidos nas projeções máximas (VP, TP, AgrP, CP) da sentença de modo livre, formando o adjunto adverbial e possibilitando o movimento livre dos verbos.

Tomando os exemplos anteriores, podemos ver as posições possíveis do advérbio de tempo “ontem” ou de dúvida/modo “provavelmente” no PB, como vemos abaixo:

- (7) a. (Ontem) O João (ontem) leu (ontem) o livro (ontem).
 b. (provavelmente) O João (provavelmente) leu (provavelmente) o livro (provavelmente).

Observando as posições assumidas pelos advérbios e o esquema padrão do processo da adjunção à projeção máxima em (8), esquematizada por Santos (2011, p. 60), verifica-se o XP2 duplicado, possibilitando a adjunção adverbial em XP1, mantendo o X' e o núcleo da projeção máxima (XP1) a esquerda:

(8):



Os exemplos em (7) demonstram que a adjunção resolve a posição estrutural dos advérbios duplicando o XP, assim, não ignora o fato de muitos dos advérbios não terem uma posição fixada na sentença no português, aloca-se em várias posições da sentença sem alterar o significado, dando conta de advérbios que precisam de uma leitura mais orientada ao objeto, como a paráfrase de (7b) possa indicar, “O João leu o jornal, não o livro”, remete a dúvida indicada pelo “provavelmente”, além de outras possibilidades das relações de dúvida entre os constituintes.

A defesa da adjunção dos advérbios e dos sintagmas adverbiais também é feita por Ernst (2004). A partir dos estudos do Programa Minimalista, ele assume que os adjuntos concatenam os advérbios e não os especificadores funcionais, porém, diferente de Costa (2004), Ernst percebe que alguns advérbios em posição de complemento do verbo movem-se para posições de especificadores. Com os AdvP adjungidos às projeções funcionais, a sua ordem é regulada por princípios semânticos, tais princípios estão distribuídos de acordo com a classificação dos adjuntos adverbiais (ERNST, 2004):

- (9) a) Sentenciais predicacionais:
- (i) Ato de fala: *frankly* - francamente, *maybe* - talvez, *luckily* – felizmente, *obviously* – obviamente.
 - (ii) Orientado ao sujeito: *deliberately* - deliberadamente, *stupidly* - estupidamente.
 - (iii) Comparativo: *similarly* - similarmente
 - (iv) Evento interno: *tightly* - firmemente, *partially* – parcialmente.
- b) Domínio: *mathematically* - matematicamente, *chemically* - quimicamente
- c) Participante: *on the wall* - na parede, *with a bowl* - com uma bacia, *for his aunt* – para sua tia.
- d) Funcional:
- (i) Relacionado com o Tempo: *now* - agora, *for a minute* - por um minuto, *still* - ainda.
 - (ii) quantitativo: *frequently* - frequentemente, *again* - novamente, *precisely* - precisamente.
 - (iii) Foco: *even* - mesmo, *just* - apenas, *only* - só
 - (iv) negativo: não
 - (v) Relações casuais: *purpose* - finalidade, *causal* - causalidade, *conditional* - condicional, etc.

Essa classificação demonstra que os advérbios são determinados por suas relações de escopo e expressas na adjunção às categorias mais importantes da sentença (TEIXEIRA, 2015). Essa relação do escopo seleciona a proposição ou o evento, de acordo com o que Ernst define como FEO(Fact-Event-Object), que são as áreas relacionadas com a ordem estrutural da sentença, esse ordenamento é delineado em níveis hierárquicos que, segundo Teixeira (2015), compreende: orientado a fala > fato > proposição > evento > evento especificado. Além dos níveis correspondentes aos tipos de advérbios ordenados: orientados para o discurso > avaliativos > modais > evidenciais > orientados para o sujeito > modo. Ernst demonstra a relação entre a FEO e as categorias sintáticas, demonstrando os locais nesses níveis:

- (10) [Ato de fala [Proposição [Evento [Evento-Interno V]]]]
 CP IP VP? VP

Segundo Teixeira (2015), Cinque (2004a) observa que a adjunção livre não diz nada a respeito das classes de advérbios que se manifestam uniformemente nas línguas naturais, já que há muitas noções no conceitual-intencional que poderiam ser gramaticalizadas, além de as noções serem distinções funcionais. Cinque (2004 *apud* TEIXEIRA, 2015) alega que a adjunção não pode fornecer evidências sobre a ordem relativa de um advérbio e um verbo ou um advérbio e um complemento da sentença.

Essas noções de análises sobre os advérbios serem adjungidos às projeções máximas não são unanimidades dentro das pesquisas gerativistas, como visto em linhas gerais acima. As pesquisas iniciais se preocuparam com as questões sintáticas e semânticas do advérbio, além da sua natureza, a relação de circunstancialidade ou modificador do verbo e, de maneira mais detalhista em termos de projeções funcionais. Cada pesquisador averigua a partir de evidências do comportamento e das relações com outros itens as suas características morfossintáticas, sintáticas e semânticas para estabelecer uma leitura desse item que sempre suscita questionamentos. Uma leitura do funcionamento do advérbio de pesquisa dentro do gerativismo, intitulada cartografia, evidencia alguns dos problemas da adjunção, como a dificuldade em dar conta dos advérbios transportáveis na sentença, além de determinar outras concepções a respeito do advérbio. Entretanto, Ernst (2004) quanto Costa (1996, 2004) trazem evidências para a flexibilidade da tese de Cinque (1999, 2004), já que o AdvP se relaciona de forma modificadora entre o item modificado e o advérbio, com possibilidades de várias interpretações de acordo com a sua ordem e tipo semântico apropriado.

2.3.2 Jackendoff (1972)

As propostas de estudos sobre os advérbios, à época da Teoria Padrão nos anos 60 e 70, se fixaram em tratá-lo como uma espécie de adjetivo por, supostamente, ambos compartilharem as mesmas regras transformacionais, sem se atentar sobre uma teoria específica que desse conta realmente dos advérbios. Jackendoff (1972) afirma que a falta de mecanismos teóricos de análise e descrição forçou os pesquisadores a subdividir esta categoria em uma infinidade de classes. Para contornar isso, dá-se início a uma abordagem mais séria, atribuindo uma distribuição sintática dividindo o advérbio de acordo com a sua posição na estrutura das sentenças na língua inglesa, possibilitando uma leitura mais semântica desses elementos. Nessa distribuição, os advérbios terminados em *-mente* (*-ly*, em inglês) são distribuídos em três tipos básicos: (i) posição inicial; (ii) posição final sem pausa

intervindo;(iii) posição auxiliar entre o sujeito e o verbo principal. Assim o linguista distingue diferentes classes de advérbios, observando as combinações nestas três posições básicas.

(11) *Cleverly–Clumsily*¹⁰

a. John *cleverly / clumsily* dropped his cup of coffee.

O João *inteligentemente/desastradamente* deixou cair o café.

b. *Cleverly /Clumsily* (,) John dropped his cup of coffee.

Inteligentemente/Desastradamente(,) o João deixou cair o café.

c. John dropped his cup of coffee *cleverly / clumsily*.

O João deixou cair o café *(,) *inteligentemente/desastradamente*.

A segunda classe dos advérbios ocorre em todas as posições sem modificar o significado (itens dessa classe no inglês: “quickly”, “slowly”, “reluctantly”, “sadly”, “quietly”, “indolently”, “frequently”, “immediately”, “often”, “soon”; em português: “rapidamente”, “lentamente”, “relutantemente”, “infelizmente”, “silenciosamente”, “indolentemente”, “frequentemente”, “imediatamente”, “muitas vezes”, “logo”).

A terceira classe ocorre na posição inicial, na posição auxiliar e posição final com pausa, são os advérbios pertencentes a essa classe: *evidently* e *probably*.

(12) *evidently / probably*

a. *Evidently/probably*, Horatio has lost his mind. (ibidem, p. 49)

Evidentemente/Provavelmente, Horácio tem perdido sua mente.

“Evidentemente/Provavelmente, Horácio perdeu a cabeça”.

b. Horatio has *evidently/probably* lost his mind.

Horatio tem *evidentemente,/provavelmente* perdido sua mente.

“Horácio *evidentemente/provavelmente* perdeu a cabeça”

c. *Horatio has lost his mind *evidently/probably*.

*Horácio tem perdido sua mente *evidentemente/ provavelmente*.

10. 10 No PE, Gonzaga (1997) afirma que não há possibilidades de ocorrer em posição final e inicial o advérbio sem pausas, como ocorre nos exemplos em (5), diferente do PE, o PB permite maior liberdade desses advérbios sem mudança de significado.

d. Horatio has lost his mind, evidently/probably.

Horácio tem perdeu sua mente, evidentemente/provavelmente.

“Horácio perdeu a cabeça, evidentemente/provavelmente”

A quarta classe se caracteriza por se posicionar apenas no final das sentenças e na posição de auxiliar.

(13) Completely / Easily

a. **Completely/easily* Stanley ate his wheaties.

**completamente/Facilmente* S comer seu cereal.

b. *Stanley completely/easily* ate his wheaties.

S *completamente/facilmente* comer seu cereal.

"S *completamente/facilmente* comeu o cereal"

c. Stanley ate his wheaties *completely/easily*.

S comer seu cereal *completamente/facilmente*.

"S comeu o seu cereal *completamente/facilmente*"

A quinta classe ocorre com advérbios posicionados no final, como é o caso dos advérbios “well”, “less”, “before”, “early”, “fast”, “home”, “slow”, “terribly”, “lengthwise”, “indoors” e “downtown”. Porém, em português há ocorrência de posição pós-verbal em muitos casos;

14) Well

a. **Well* Sam did his work.

*Bem Sam fez seu trabalho

b. *Sam *well* did his work.

Sam *bem* fez seu trabalho.

c. Sam did his work *well*.

Sam fez seu trabalho *bem*.

"Sam fez o trabalho *bem*"

A sexta classe e última listada por Jackendoff (1972) apenas ocorre em posição auxiliar, a exemplo dos advérbios merely, truly e simply.

- (15) a. Albert is merely/truly/simply being a fool.
 A é meramente/realmente/simplesmente sendo um tolo
 “A é meramente/realmente/simplesmente um tolo”
- b. *Merely/Truly/Simply is Albert being a fool.
 * Meramente/realmente/simplesmente é A sendo um tolo.
- c. *Albert is being a fool merely/truly/simply.
 *A é sendo um tolo meramente/realmente/simplesmente.

Essas classificações são estabelecidas muito mais por critérios semânticos do que sintáticos. Com os estudos progredindo na Teoria Gerativa, as pesquisas evoluíram da Teoria Padrão para a P&P e a distinção entre os advérbios modificadores de VP e os de sentença não eram mais suficientes, tendo em vista que, o advérbio modifica estruturas sentenciais, sintagmáticas e lexicais (SANT’ANA, 2010). A importância do trabalho de Jackendoff (1972) se fundamenta tanto pela análise considerável de advérbios do inglês, quanto por iniciar uma abordagem a partir desse item, considerando as suas propriedades semânticas e sintáticas de acordo com a sua posição nas estruturas em que ocorrem.

2.3.3 Pollock (1989)

A proposta de Pollock (1989), diferente da proposta de Jackendoff (1972), sistematiza as diferenças entre o francês e o inglês a partir de diferentes fatores, como a sintaxe da frase, da negação, da dúvida, dos quantificadores flutuantes, da relação dos movimentos do verbo, a Gramática Universal (UG) e a estruturação do IP nessas línguas. Apresenta estudos inovadores em relação à posição de constituintes, afirmando que o advérbio possui posições fixas na sentença, mas com diferentes posições diante dos verbos lexicais em francês e dos verbos auxiliares em inglês.

Com uma abordagem voltada para as projeções em IP, Pollock (1989) oferece argumentos empíricos sobre a Flexão 11 (INFL(Inflexion)) não ser um constituinte que possui dois traços ([±Tense, ±Agr]), em vez disso, cada um desses constituintes se caracterizam como núcleos de projeções máximas. Pollock pretendia resolver as questões destes movimentos diferenciados entre essas duas línguas, dentre essas diferenças, de acordo com Kenedy (2013, p. 203), indicava “que o sintagma de tempo comportava, na verdade, mais

do que somente a expressão de tempo verbal. Ele dava conta também da negação, da concordância, aspecto, dentre outras noções funcionais”.

Ao demonstrar o movimento dos verbos, advérbios como *often/souvent*, *seldom/rarement* seriam modificadores do VP na estrutura de adjunção e suas posições seriam fixas justamente por o verbo alçar para buscar seus traços no IP, agora cindido em duas categorias ([TP, AgrP]). Como vemos abaixo, a posição do advérbio (V Adv) em francês é oposta a posição ocupada pelo advérbio equivalente em inglês (que possui verbos lexicais fixos – Adv V):

- (16) a. Jean embrasse *souvent* Marie. (ibidem, p. 367)
 b. *John kisses *often* Mary.
 c. *Jean *souvent* embrasse Marie.
 d. John *often* kisses Mary.

Os exemplos (16a e 16b) demonstram que os advérbios em questão ocupam, no francês, uma posição pós-verbal e em inglês uma posição pré-verbal, em outras posições (16b, 16c) a sentença se torna agramatical. A posição dos advérbios *souvent* e *often* ocorrem antes do VP ou posicionados no VP inicial nas estruturas-D, para resultar nas ordens de superfície [V Adv] no francês após o movimento do verbo; e [Adv V] sem movimento de verbos lexicais no inglês. O comportamento desses advérbios no PB demonstram que essa língua torna opcional a posição do advérbio sem sofrer agramaticalidade ([Adv V] [V Adv]) demonstrando que os verbos se movimentam para AgrP e TP.

As evidências são verificadas diante de verbo auxiliar na estrutura superficial. Todos os verbos auxiliares tanto do inglês quanto do francês se movimentam ([Aux Adv V]) para INFL. Porém, no português, é evidenciado uma possível restrição do advérbio ao se posicionar entre o verbo auxiliar e o verbo principal.

Pollock (1989)¹¹ buscou a ordem dos advérbios para demonstrar que no francês os verbos se movem para Infl na posição pre-verbal; e em inglês, apenas os verbos auxiliares (have, be) se movem para Infl. O PB parece aceitar os movimentos de ambas as posições das línguas descritas. Além de propor a cisão do IP para compor dois traços importantes na Gramática Gerativa, o autor conclui que os advérbios se inserem na sentença por adjunções

11. 11 O trabalho de Belletti (1990) segue os pressupostos de Pollock (1989) na análise do movimento sistemático do verbo em Italiano e a ordenação dos advérbios na sentença.

graças a essas regras de movimento do verbo, caracterizando o advérbio não possuidor de movimento nestas línguas por serem gerados na base do léxico.

2.3.4 Laenzlinger (1998)

Desenvolvendo a sua teoria sobre a sintaxe dos advérbios, Laenzlinger (1998) afirma, inicialmente, que os advérbios são elementos que podem ocupar várias posições na sentença em relação ao sujeito, verbo e complemento, mas diferem as posições lineares de um idioma para outro. Ele verifica que alguns advérbios possuem multi-posições como os advérbios modais em francês. O autor também verifica as ordens no inglês e no francês, atestando o que afirmou Pollock (1989). Exemplificando com o advérbio *souvent* e *often* para demonstrar que os advérbios ocupam varias posições graças às propriedades de iniciar na sentença e multi-posições, além de ser afetado pela ocorrência dos verbos-argumento. Vejamos o exemplo do próprio autor (*ibidem*, p. 4).

- (17) a. (Souvent,) Jean (, souvent,) a (souvent) embrassé (souvent) Marie (souvent)
 b. (often) John (,often,) has (often) kissed (often) Marie (often).
 "John often kissed Mary"¹²

Os exemplos acima, também discutido em Pollock (1989), demonstram as posições possíveis que os Adv podem ocupar sem causar agramaticalidade, corroborando a sua hipótese de que estes itens se apresentam mais livres na sentença do que os itens verbais. Como tentativa de responder se a posição sintática de um advérbio é determinada por distribuição, Laenzlinger (1998) propõe descrever alguns conceitos presentes na literatura, elaborando sem pretensões de ser definitiva uma classificação, mas deixa claro que o seu objetivo é definir as condições formais de licenciamento das distribuições dos advérbios.

O autor observa que tradicionalmente os advérbios são divididos em dois grupos, advérbios sentenciais e advérbios de VP, com características em modificar uma sentença inteira e modificar um verbo na sentença. Há mais dois grupos nomeados como advérbios frasais que podem modificar uma série de frases e os advérbios adjetivais (a exemplo de “très” - “very” - “muito”) que normalmente modifica na sentença, adjetivos e outros advérbios.

12. João beijou frequentemente Maria.

Resumindo as condições de distribuição dos Adv-VP e dos S-advérbios, o autor demonstra uma série de definições para a sua teoria sintática dos advérbios, como vemos abaixo a sua classificação:

• **Advérbios Sentenciais (S-Adv):**

1. Advérbios Pragmáticos:

- a) (a) Avaliativos: felizmente - heureusement - fortunately;
- b) (b) Conjuntivos: porém - cependant - however;
- c) (c) Formais: precisamente - précisément - precisely;
- d) (d) Orientados: francamente - franchement - frankly.

2. Advérbios de Domínio:

- a) Logicamente - logiquement - logically.

3. Advérbios Modais:

- a) Provavelmente - probablement - probably.

4. Orientados ao sujeito:

- a) Passivo: deliberadamente - intentionnellement - deliberately;
- b) Orientado ao Agente: violentamente - violemment - violently.

5. Advérbios de Eventos:

- a) Tempo: ontem - hier - yesterday;
- b) Local: ici - here.

6. Advérbios Aspectuais

- a) Frequência: fréquemment - frequently;
- b) Movimento: lentamente - lentement - slowly.

• **Advérbios de VP:**

1. Advérbios Aspectuais:

- a) Positivo - Assertivo: sempre - toujours - always, verdadeiramente - vraiment - truly;
- b) Quantificador: muito - beaucoup - much, suficiente - assez - enough;
- c) Grau: quase (praticamente) - presque - almost, dificilmente - à peine - hardly.

2. Circunstancial (modo, instrumento, resultado etc.):

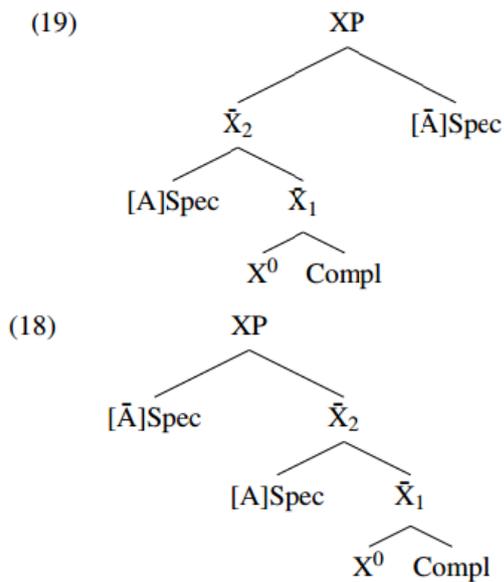
- a) Orientado ao Verbo: meticolosamente - pémblement - painstakingly, correcteme - orrectly;
- b) Orientado ao Objeto: totalmente - entièrement - entirely, fortemente - loudement - heavily.

O desenvolvimento terminológico, segundo Laenzlinger (1998), se desenvolve a partir de Jackendoff (1972), Belletti (1977) e Ernst (1984). Essas distinções foram observadas a partir das várias posições assumidas na sentença pelos advérbios nos trabalhos desses pesquisadores, em especial os Advérbios de VP, segundo o autor, esses advérbios seguem determinando o verbo na sentença. Vale salientar que ao definir a constituição lexical, categorial e funcional, argumenta que o advérbio é uma categoria que não se enquadra na tradicional classificação dos traços [V N] [+ -]. Apesar de ser uma categoria gramatical, a maioria dos itens não possui a capacidade de subcategorizar, além disso, é capaz de projetar uma categoria máxima.

Na questão semântica, os critérios que estabelecem um princípio formal que legitima a distribuição do advérbio em Laenzlinger, passa, segundo Gonzaga (1997), por um critério-Adv, que é uma particularização do Critério Generalizado de Sportiche (1992). Este critério define a camada categorial e funcional dos advérbios na sentença.

Além desse critério, Laenzlinger assume os advérbios como especificadores (Spec) das categorias funcionais na mesma linha de Cinque (1999), ao assumir que eles são operadores funcionais, não adjuntos adverbias. Já que os adjuntos adverbias, mesmo se desenvolvendo nos estudos do Princípios e Parâmetros (P&P), contêm sensíveis diferenças teóricas e empíricas. Essa opção proposta de utilizar os advérbios em Spec de um núcleo funcional lhe faz propor duas operações para dar conta da substituição dos adjuntos adverbias. Consiste em multiplicar as projeções funcionais e, quando for pertinente, o enriquecimento do modelo X-barra.

Ao propor um novo modelo à X-barra que projetasse a estrutura clássica sem o adjunto (Especificador, Núcleo e Complemento), mas a duplicação da projeção funcional e o enriquecer o X-esquema (*X-schema*), o autor adota a opção propondo no modelo uma estrutura que permite, no máximo, dois especificadores por projeção (*A-specifier*). Nos exemplos (18) e (19) abaixo temos duas expressões abstratas dessa duplicação dos especificadores.



Nestes esquemas projetados por Laenzlinger (1998, p 75), a projeção máxima de (18) recebe um especificador não argumental em irmandade com o X₂. Há especificadores argumentais em irmandade com as projeções intermediárias de X₁ que domina o X₀ e o Compl(emento) na estrutura (19). O especificador [A]Spec é licenciado graças às características da projeção do núcleo, possuindo características dos traços que licenciam o WH, Top, Foc e Neg. Já a realização do especificador [A]Spec argumental é licenciada graças à presença de recursos no núcleo que possibilitam as funcionalidades dos argumentos no Spec. Essas características argumentais e não argumentais dos especificadores podem ser realizadas na mesma projeção do núcleo.

1. Quantificadores: atribuem valor quantitativo a uma variável, são normalmente aspectuais, negativos, quantitativos e focalizadores;...
2. Qualificador: advérbios que atribuem qualificação a uma variável, normalmente são temporais, locais e de modo.

De acordo com Laenzlinger (1998), O Critério-Adv faz parte de um sistema de critérios que licenciam os elementos quantificadores, assim como há o Critério de Sentenças-h, Critério-Neg para as negações, Critério-Foc para constituintes Focalizadores. Para os qualificadores, os elementos precisam ter seus traços checados para serem licenciados, como é o caso dos traços de Caso, por exemplo. Essa relação lexical dos quantificadores e os Critérios-Adv configuram o Spec-núcleo, mas suas aplicações se diferem, pois um Critério é satisfeito por um núcleo (elemento lexical) no final da derivação, já a checagem de traços na

maioria das vezes, por questões morfológicas, é satisfeita por qualquer elemento no percurso da derivação (SANTOS, 2011).

Seguindo uma proposta de hierarquia para os advérbios, Laenzlinger (1998) estabelece a partir das relações de checagem de traços e das projeções funcionais para distinguir a realização de cada advérbio na sentença, levando em conta, não apenas a semântica, também a sua distribuição (licenciamento) com os elementos propostos em (12) resumidos em Advérbios de Sentença (modo, Orientados ao sujeito, tempo e localização, frequência e movimento) e Advérbios de VP (orientados ao verbo, orientados ao sujeito, aspectuais). A hierarquia interna ao advérbio é mapeada dentro das projeções na sentença, a exemplo de “Provavelmente, estes soldados não comem bem recentemente” em francês (ibidem, p. 90):

- (20) a. [*CP Adv modal [TP Adv evento [AspP/IP Adv frequência [AspP/VP Adv quantificador [VP Adv modo]]]]]*
 b. [*CP [AgrsP cesmilitaires n’ont [NegP pas [TP récemment [VP bien mangé]]]]]*]

A observação de Laenzlinger (1998) sobre a distribuição percorre diversos fenômenos distintos que demonstram a dependência sintática sobre esses fenômenos (além dos citados, ele traz o Movimento do Verbo; a Extraposições no NP; a Distribuição dos advérbios no CP). O autor conclui afirmando que buscou demonstrar as correlações entre a forma lógica dos advérbios e a sua distribuição na estrutura hierárquica, propondo, assim, mecanismos reguladores dessa distribuição, tais como operadores que ocupam posições de especificadores, como o Critério-Adv para os advérbios distribuídos como quantificadores e advérbios qualificativos que são sujeitos a Checagem de traços. Laenzlinger se figura entre os teóricos cartográficos que propõem o licenciamento dos advérbios tanto no Spec-X quanto no Spec-P. Como vimos, Laenzlinger (1998) advoga em favor de os advérbios se figurarem como especificadores e não adjuntos. Apesar de os estudos cartográficos apresentarem evidências para o advérbio figurar como Spec-X ou Spec-XP, não violando os princípios assimétricos proposto por Kayne (1994).

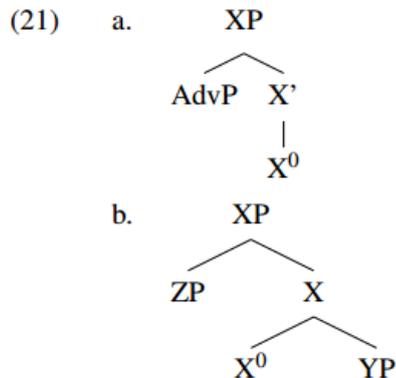
2.3.5 Cinque, 1999, 2004, 2006

Com uma perspectiva que vai de encontro a adjunção dos advérbios na sentença, Cinque (1999, 2004a, 2006) questiona a definição de advérbios como núcleos pertencentes a projeção máxima do VP, ou de outra projeção estendida dominada pelo VP, além de complemento do VP. Para Cinque (1999), os advérbios não bloqueiam o movimento dos

núcleos verbais, além de verificar que muitos advérbios se movimentam como topicalizadores e focalizadores que são checados nas projeções máximas (XP), menos para as projeções de núcleo (X⁰).

Para Cinque (1999), os itens lexicais denominados advérbios são especificadores de diferentes projeções máximas, não adjuntos, e são distribuídos de forma fixa de acordo com o que está determinado na Gramática Universal (GU). Essa análise proporciona uma compreensão mais detalhada do posicionamento sintático dos sintagmas adverbiais a partir do que Cinque (1999, 2006) conceitua como Hierarquia Linear Universal (HLU) das projeções funcionais. Pretendendo corroborar a hipótese de que a GU não permite variações em número e tipos de projeções funcionais, sendo assim, não existem diferentes projeções funcionais nas línguas, mesmo elas demonstrando evidências de inúmeras restrições e movimentos nas estruturas diferenciadas nas línguas naturais.

Sobre o Sintagma AdvP e o posicionamento sintático dos advérbios na estrutura sintática da linguagem, Cinque não despreza as estruturas de Pollock (1989) e tampouco o LCA de Kayne (1994), exemplificados abaixo:



Na representação (21a), Santos explica que a estrutura de um XP com uma projeção funcional que faz *merge* com a sua projeção intermediária X' e os sintagmas adverbiais ocupando a posição de Spec em uma relação *agree* entre o SAdvP e o núcleo funcional da sentença, demonstra com fidelidade a estrutura arbórea de Cinque (1999), o mesmo pode ser replicado em (21b). Dentro dessas estruturas, poderemos afirmar que Cinque estabelece uma ordem dos principais AdvP nas sentenças (aplicadas juntas nas línguas românicas e algumas línguas Crioulas e Asiáticas), tal ordem se trata de uma hierarquia universal para o sintagma adverbial.

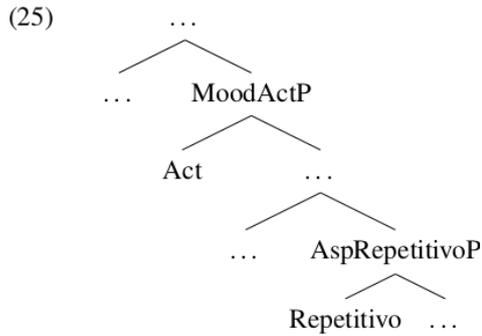
A Hierarquia dos AdvP Funcionais é observada a partir de algumas restrições impostas entre os advérbios na mesma sentença, como nos mostra o autor no italiano, o advérbio negativo *non*, normalmente é antecedido pelo advérbio *solitamente* (normalmente), e *mica* antecede o advérbio *già* (já), vejamos o exemplo:

- (22) a. Alle due, Gianni non ha *solitamente mica mangiato*, ancora.
 ‘Às duas, G. não tem normalmente mica comido, ainda’
 b. *Alle due, Gianni non ha *mica solitamente mangiato*, ancora.
 “Às duas, G. tem não normalmente comido ainda”.
- (23) a. Non hanno *mica gia* chiamato, che io sappia.
 ‘Não ter já mica chamado que eu sei’.

Com essas observações, Cinque percebe que há uma hierarquia que deve ser levada em consideração, pois o resultado de [*solitamente > mica > gia*] formalizado com outros advérbios, pode-se chegar a um esquema das ordens hierárquicas definidas na GU. Mas à frente, Cinque (1999) demonstra o esquema para outros advérbios, por exemplo: [*solitamente > mica > gia > piu > sempre > completamente > tutto > bene*] e [*francamente > fortunatamente > evidentemente > probabilmente > ora > forse > intelligentemente*]. Diante dessas evidências, ele divide os tipos de Advérbios Baixos (Pré-Verbais *usually, again, often, already, soon*, etc) Advérbios Altos (Sentenciais – *frankly, unfortunately, apparently, probably*, etc.) de acordo com a posição. Diante dessas evidências, Cinque (1999, p. 106) propõe o esquema abaixo:

- (24) Mood (ato de fala) > Mood (avaliação) > Mood (evidencial) > Mod (epistêmico) > Tempo (passado) > Tempo (futuro) > Mood (irrealista) > Mod (alético de necessidade) > Mod (possibilidade) > Aspecto (habitual) > Aspecto (repetitivo) > Aspecto (frequentativo) > Mod (volicional) > Aspecto (acelerativo) > Aspecto (continuativo) > Aspecto (perfectivo) > Aspecto (retrospectivo) > Aspecto (aproximativo) > Aspecto (durativo) > Aspecto (genérico/progressivo) > Aspecto (prospectivo) > Aspecto (completivo I) > Aspecto (completivo) > Voz > Aspecto (acelerativo(II)) > Aspecto (repetitivo(II)) > Aspecto (frequentativo) > Aspecto (completivo).

Do esquema acima, a estrutura resultante que representa a proposta pode ser demonstrada com algumas projeções internas hierarquizadas por Cinque, como podemos ver abaixo, o MoodActP representando o Modo Ato de Fala acima do AspRepetitivoP que representa o Aspecto Repetitivo:



Essas são as projeções funcionais na arquitetura sintática das línguas naturais proposto por Cinque (1999). Esses elementos estão relacionados com as categorias funcionais de Tempo, Aspecto, Modo, Voz e Número, e estão ligados às categorias verbais. Sobre as projeções, elas estão divididas, como indicado anteriormente, em sintagmas mais baixos ou mais altos. A ordem dos advérbios baixos relacionados, equivalentes em PB, delimitadas por Santos (2011) e Sant’ana (2010), são geralmente > Neg> já/ainda> mais/ainda> sempre/jamais> completamente/parcialmente> tudo/nada> bem/mal. Já os advérbios altos são francamente> felizmente> evidentemente> provavelmente> agora> talvez> inteligentemente. As evidências utilizadas por Cinque (1999) foram extraídas de diversas línguas europeias, asiáticas e crioulas, assim, a hierarquia dos especificadores postula uma ordem universal para os funcionais.

Os AdvP mais baixos ocorrem na posição mais baixa da sentença, de acordo com Cinque (1999), em um espaço delimitado, de um lado a posição mais à esquerda que um participio passado ocuparia, e no outro lado um complemento ou sujeito do participio passado. Esses itens pertencentes a esse espaço obedecem a uma ordem rígida, como solitamente “geralmente” e mica “não”. Já os AdvP (sentenciais) altos englobam a classe dos advérbios propostos por Jackendoff (1972), os advérbios orientados para o sujeito e os orientados para o falante. De acordo com a hierarquia, os advérbios orientados para o falante (provavelmente) precedem (>) os advérbios orientados ao sujeito (inteligentemente), deixando evidências para que os advérbios altos não formem uma classe homogênea.

Os AdvP que ocorrem em posição final (VP-Final), menos os AdvP (sentencias) só ocupam a posição final mediante pausa entoacional:

- (26) a. *Non posso sopportare neanche Carlo onestamente
 b. Non posso sopportare neanche Carlo, onestamente
 Não posso suportar mesmo Carlos, honestamente

Os AdvP que ocorrem na posição pós-complemento com ou sem pausa entonacional, carregando a marca de foco e obedecendo à mesma ordem do espaço pré-VP:

- (27) a. Gianni non vince le sue partite già più sempre bene;
 b. *Gianni non vince le sue partite già bene più sempre.
 G. não vence as suas partidas já mais sempre bem.

Santos (2011) e Sant'ana (2010) descrevem algumas projeções funcionais em relação ao português, iremos reproduzir tais descrições resumidamente para evidenciar melhor os tipos de projeções encontradas em Cinque (1999, 2006):

Modalidade ato de fala – *francamente, honestamente, sinceramente, etc.*: Marcam a força ilocucionária da sentença; posição mais alta no espaço do IP (categoria funcional de flexão).

Modalidade avaliativo – *(in)felizmente, lamentavelmente, surpreendentemente, estranhamente, inesperadamente*: Expressa a avaliação (positiva, negativa, duvidosa, etc.) sobre o estado de coisas descritos na proposição; Não afeta o valor de verdade da proposição.

Modalidade evidencial – *supostamente, claramente, evidentemente, obviamente etc.*: expressa a evidência auditiva, sensorial, relativa evidencial (revelação em um sonho ou premonição) experiencial etc.; possui uma marcação bem elaborada e complexa, podendo marcar diferentes distinções avaliativas de acordo com a concordância com o testemunho visual do falante sobre o estado, evento ou processo descritos na asserção.

Modais epistêmicos – *provavelmente, presumivelmente, supostamente*: Expressa o grau de confiança sobre a proposição, ou seja, que expressa verdades possíveis ou necessárias baseada na informação.

T(passado) e T(futuro) – *depois e antes*: Indica a anterioridade ou posterioridade de um evento em relação a outro expresso ou não na sentença.

Modalidade Irrealista – *talvez e provavelmente*: Indica a incerteza sobre a verdade da proposição ou a incerteza sobre o saber do destinatário. Expressando ações não concretas, não verídicas, algo passível de contestação.

Modais alético – *(des)necessariamente* e *possivelmente*: indicam as verdades necessárias, tanto as proposições sempre verdadeiras quanto as proposições verdadeiras em algum cenário.

Aspecto habitual – usualmente, habitualmente, geralmente, regularmente etc.: Descreve uma situação característica de um certo período de tempo; ligados a categorial verbal de Aspecto.

Aspecto repetitivo-frequentativo I e II – Sem um correspondente efetiva no PB (novamente > de novo): Repetição de uma ocasião ou várias ocasiões.

Modal de volição – Sem advérbios no PB relacionados: Advérbios orientados pelo sujeito; modais de raiz (*root modals*) mais próximos ao radical do item lexical.

Aspecto acelerativo – *rapidamente*: a depender de quais advérbios estejam relacionados a esse núcleo, pode tomar escopo sobre o evento ou sobre processos de acordo com a posição na estrutura da sentença que venha ocupar.

T(anterior) – *Já*: indica prioridade temporal a um tempo marcado como referência; nos períodos compostos, obriga uma leitura prioritária para o evento marcado na sentença em que se encontra.

Aspecto continuativo – *ainda*: marca a duração de um evento iniciado anteriormente ao ato de fala.

Aspecto perfectivo / imperfectivo – *sempre*: Sem verificação satisfatória da relação entre o advérbio e o núcleo.

Aspectos retrospectivo e aproximativo – *recentemente*, *ultimamente*, *imediatamente*, *subitamente*: O retrospectivo se relaciona a um evento ocorrido um pouco antes da referência de tempo; o aproximativo se relaciona a outra noção de um evento que segue outro.

Aspecto durativo – *longamente* e *brevemente*: Indica uma ação durável por um certo tempo.

Aspecto genérico/progressivo – *caracteristicamente*, *inerentemente* e *tipicamente*: uma propriedade inerente de um objeto manifestada ou não (por exemplo: (Tipicamente) Este sapato (tipicamente) dura (tipicamente) o mês inteiro (,tipicamente.); não há clareza sobre quais classes de advérbios corresponda a esse núcleo aspectual.

Aspecto prospectivo – *quase* e *imediatamente*: indica um ponto antes do início de um evento.

Aspecto completivo – *completamente*: demonstra que um objeto foi afetado por uma ação ou por uma característica.

Voz – inúmeros advérbios em PB ocupando essa posição, mas sem correspondência direta: Classe que ocupa a posição de especificador de VoiceP, mesmo que a relação não seja evidente; equivalente à noção de advérbio de Modo da GT.

Aspecto acelerativo II – *Quickly* e *rapidly*: no inglês ocorrem em duas posições diferentes, mas com interpretações distintas.

Aspecto completivo II – Sem correspondência com o PB: No inglês, *completely* (completamente) aparece em duas posições diferentes, mas com interpretações distintas.

Esses conceitos demonstram que o ordenamento dos AdvP segue o posicionamento rígido hierárquico das projeções funcionais da arquitetura na sintaxe. Cinque (2006, p. 12) atualiza a HLU, sugerindo uma melhor correspondência com as articulações e a rigidez das classes dos AdvP. Com isso, pensar em mobilidade livre dos advérbios é uma aparente observação, pois os mesmos já estão definidos na GU, apenas no processo da aquisição, as posições irão ser preenchidas pelos itens lexicais apropriados, a começar pelos advérbios mais altos pré-verbais (CINQUE, 1999).

Os estudos de Cinque (1999, 2006) preveem uma universalidade das projeções funcionais dos advérbios, observando a ordem como cada elemento é posicionado na estrutura sintática, prever até uma facilidade na aquisição da linguagem por essas projeções funcionais já estarem presentes na GU. O empreendimento Cartográfico nos fornece possibilidades em fornecer perspectivas no tratamento dos advérbios na estrutura das sentenças, tanto em ordenamento quanto em traços informacionais mais detalhados.

2.4 A DELIMITAÇÃO DOS ADVÉRBIOS

As questões relacionadas aos advérbios, desde os gramáticos aos estudos mais científicos da linguística com a teoria gerativa, estabelecem observações sobre o comportamento os advérbios com propósitos de atingir uma descrição mais exaustiva e explicações mais claras sobre os elementos dessa categoria. Como o objetivo é uma formalização mais próxima do Minimalismo, como vimos, temos Ernst (2004) e Costa (1996) na perspectiva de adjunção e; Cinque (1999, 2004, 2006) e Laenzlinger (1998) fornecendo uma perspectiva pró advérbio-Spec nas posições pré-VP, além de uma estrutura sintática mais próxima de Kayne (1994).

O que pode escapar dessas concepções são os movimentos para a checagem de traços que postulam movimentos e concebem ao advérbio projeções funcionais de acordo com a sua posição, tornando uma solução nada trivial para as modelagens computacionais e para as

formalizações de *parsers*. Mesmo que há a divisão de advérbios como falsos ou verdadeiros, ainda assim, o comportamento determinístico de qualquer advérbio, quando não lhe escapa a função adverbial, pede uma classificação para preenchimento dos nós terminais exigidos pela assimetria e pelo modelo computacional. Por esses motivos, classificar os advérbios na sua base lexical é sempre obrigatória, mesmo que delicado para qualquer postulação gerativo e linguístico computacional. Por isto, os estudos de Cinque se mostram mais adequados por conceber naturezas e traços funcionais já determinados sobre os advérbios.

Dentro da literatura gerativa, os advérbios terminados em *-mente*, por vezes, são considerados verdadeiros advérbios por ser, em sua maioria, derivados de adjetivos e terem uma maior mobilidade na sentença, além de ser considerados um subgrupo do mesmo por evidências morfológicas (BOMFIM, 1988; OLIVEIRA, 1996; OLIVEIRA, 2010; LIMA, 2006). Por não haver ainda uma pesquisa que aprofunde a classificação para os advérbios denominados Pseudos-Advérbios em Lima (2006), manteremos a classificação geral para todos os itens aqui classificados como advérbios, mas usaremos os terminados em *-mente* nos exemplos aqui vistos. Diante de inúmeras possibilidades de classificações para o advérbio, devido às questões sintáticas e semântica, veremos as relações de ordem de alguns advérbios e seus valores.

2.4.1 A ordem dos advérbios

Na gramática gerativa, há diversas propostas para explicar a distribuição dos advérbios, sejam como especificadores, sejam como adjuntos. Assim como a distribuição, as definições, as funções e as categorizações se diferenciam, em maior ou menor grau, de autor para autor, a ordem dos advérbios diverge em relação categorial e funcional. Na ordem possível de ocorrência dos advérbios na linearidade dos constituintes, Oliveira (1996, p. 71) lista essas possibilidades para o PB:

- (28) a. Adv + S + V + O
 b. S + Adv + V + O
 c. S + V + Adv + O
 d. S + V + O + Adv
 e. Adv + S + V auxiliar + V principal + O
 f. S + Adv + V auxiliar + V principal + O

- g. S + V auxiliar + Adv + V principal + O
- h. S + V auxiliar + V principal + Adv + O
- i. S + V auxiliar + V principal + O + Adv

São poucos os casos que não há a possibilidade de um advérbio estar em várias posições na sentença no PB, de acordo com o exemplo (28), porém, precisamos verificar quais são esses advérbios e por quê. Segundo Lima (2006), os advérbios terminados em *-mente*, possuem uma maior mobilidade (liberdade) que os outros (falsos-)advérbios na sentença, com um comportamento mais sensível na sintaxe, evidencia a existência de dois tipos de advérbios. Essa diferença dos advérbios terminados em *-mente* pode estar determinando este comportamento sintático mais livre, porém, essa diferenciação morfossintática, por questões de proposições e definições mais abrangentes, necessita de mais análises. O mais importante no momento é a afirmação de que nem todos os advérbios possuem as mesmas possibilidades dos terminados em *-mente* e quais seriam eles, pois se tomarmos o advérbio “sempre”, seria facilmente observado a sua mobilidade igual aos advérbios terminados em *-mente*, por exemplo. Vejamos o exemplo de Silva (2001, p. 60) em (29):

- (29) a. A Bia sempre faz o dever de casa.
- b. A Bia faz sempre o dever de casa.
- c. A Bia faz o dever de casa sempre.
- d. Sempre a Bia faz o dever de casa.

Nos exemplos acima, Silva observa que em (29a), o advérbio “sempre” está entre o sujeito e o verbo, essa posição já supõe o advérbio adjungido para T’ e o verbo localizado em T. Em (29b) o advérbio é adjungido no Asp(ecto) por seguir o verbo. Quando se encontra em posição VP-Final, como em (29c), o advérbio é adjungido à V’ e recebe uma leitura focalizada. Incluímos o (29d) para mostrar que o advérbio sempre pode se localizar no início da sentença sem provocar agramaticalidade. A partir desses exemplos e para efeito de demonstração, as estruturas das sentenças (29) devem ser, segundo a autora, como mostra (30) abaixo:

- (30) a. [TP A Bia [T' sempre [T faz [AspS t s [Asp' t v [VP t s [V' t v o dever de casa]]]]]]]
 b. [TP A Bia [T' faz [AspS t s [Asp' sempre [Asp' t v [VP t s [V' t v o dever de casa]]]]]]]
 c. [TP A Bia [T' faz [Asp' t v [VP t s [V' [V' t v o dever de casa] sempre]]]]]

Como observado anteriormente, Jackendoff (1972)¹³ defende a existência do advérbio como uma categoria lexical independente e, assim como o adjetivo superficializa entre um determinante e o nome, a distribuição dos advérbios é entre o nome (sujeito) e o verbo na língua inglesa. Cinque (1999) defende que os advérbios ocupam posições na Categoria Funcional (CF) da sentença, se hierarquizam e se dividem entre advérbios baixos (Pré-VP) e altos (escopo). Com as restrições observadas por Cinque, o ordenamento relativo fixo dos advérbios é explicado pela sua hierarquia dos especificadores visto anteriormente, através dos núcleos funcionais a partir da concordância entre o especificador e o núcleo, com isso, a projeção teria uma interpretação semântica específica para cada advérbio. Como vimos, os advérbios posicionados como especificadores são projeções de um núcleo funcional e a rigidez da ordem é especificada pela GU, além de que a assimetria modela uma estrutura sintática facilitando a explicação do comportamento assimétrico dos advérbios na sentença.

Sant'ana (2010, p. 56) explica que o argumento de Cinque postula o advérbio fixo na sentença enquanto o verbo se movimenta, demonstrando uma posição de núcleo imediatamente à esquerda e outra à direita de cada advérbio na CF, seguindo o raciocínio de Cinque, temos a seguinte forma: Tomando A e B como elementos representando os advérbios, a ordem deles são imutáveis. Com um novo elemento C representando o verbo, as ordens possíveis serão C - A - B; A - C - B; A - B - C. Qualquer outra ordem será agramatical. Porém, a língua portuguesa é um desafio para as afirmações de Cinque, como podemos ver em (31) o advérbio de Ato de Fala (francamente) precede o Avaliativo (felizmente), se houver inversão de posições, ocorre também uma possível modificação do significado, orientada para o sujeito, talvez.

- (31) a. Francamente, felizmente João serviu o bolo. (ibidem, p. 93)
 b. ??Felizmente, francamente João serviu o bolo.

13. 13 Jackendoff (1972) demonstra alguns advérbios ocuparem as três posições com alteração de significado na mesma sentença, colaborando com a evidência sobre a dicotomia advérbios verdadeiros e advérbios falsos não ser sustentada apenas com o argumento da mobilidade diferenciada dos advérbios terminados em -mente em relação aos outros tipos não terminados em -mente, também o comportamento semântico pode ser mais uma evidência para essa possibilidade de limitar e identificar com mais detalhes os tipos de advérbios e os que não são, de fato, pertencentes a essa categoria tão heterogênea.

Outras ordens desses advérbios são possíveis, mas priorizamos os advérbios descritos por Cinque (1999, 2006) e utilizados na análise de Sant'ana (2010) para confirmar as posições de cada classe de advérbio que serão listados posteriormente. É importante lembrar que essa ordem, até o momento, é generalizada na construção dos *parsers* nas análises das sentenças em que há presença dos advérbios. Contornando essa generalização, listaremos os advérbios ditos altos, seguido dos advérbios baixos para uma melhor formalização. De acordo com a hierarquia de Cinque e os pressupostos minimalistas, podemos organizar a partir do léxico as representações funcionais correspondentes aos tipos de advérbios.

Em relação à hierarquia de núcleos funcionais, Sant'ana (2010) confirma a sua aplicação no PB, porém, é preciso observar algumas discrepâncias, como os núcleos que projetam os Atos de fala, Avaliativos, Epistêmicos e Evidencias que são projetados para o complementizador; os advérbios de Ato de fala ocupam uma das duas posições no CP; o núcleo Epistêmico e Irrealista possuem a mesma característica semântica; não ocorrem Completivo (I) e Voz no PB; Já e Ainda podem ser atribuídos a um mesmo núcleo (Tempo contínuo/anterior) antes do Terminativo. Concluindo assim uma ordem hierárquica para as projeções funcionais do PB a partir da sua análise dos advérbios Altos e Baixos.

(32) Atos de fala(I)> Atos de fala(II)> >avaliativo > evidencial > epistêmico > alético
necessidade > habitual > repetitivo (I) > frequentativo (I) > aspecto acelerativo (I) >
tempo (anterior/contínuo) > aspecto continuativo > aspecto perfeito/imperfeito >
aspecto retrospectivo/aproximativo> aspecto durativo > aspecto genérico/progressivo >
aspecto acelerativo (II) > aspecto repetitivo (II) > aspecto frequentativo (II) >
completivo (II).

Listaremos esta hierarquia de (32) nos exemplos que demonstram no PB as suas ocorrências de acordo com a hierarquia das projeções dos núcleos funcionais estabelecidos por Santa'ana (2010).

2.4.2 Advérbios Altos

Classe dos advérbios posicionados naturalmente no início da sentença, como podemos ver no exemplo (32) abaixo, retirado de Sant'ana (2010, p. 98).

- (33) a. Ato de Fala Francamente, João não deveria sair à noite.
 b. Avaliativo Felizmente João comeu a carne.
 c. Evidencial Evidentemente João assinou o documento.
 d. Epistêmico Supostamente João comeu a carne.
 e. Tempo Então João comerá a carne¹⁴.
 f. Aléticos Necessariamente João comerá a carne.

No (33a) o advérbio é o mais alto da hierarquia, se posiciona mais à esquerda da sentença e indica a impossibilidade de serem antecipados por outros constituintes sintáticos, além de ser separados da sentença por uma pausa prosódica ao marcar a força ilocucionária da sentença. Os advérbios de ato de fala estão localizados no PB, segundo Sant'ana, no complementizador, não na camada funcional. Os outros advérbios iniciais mostrados em (33) ocupam as posições nas camadas mais baixas no complementizador das sentenças. Em (34), listamos os advérbios ato de fala que estão mais à esquerda da sentença, demonstrando a posição fixa no complementizador da sentença.

- (34) a. Francamente, João deveria servir o bolo. (ibidem, p. 113)
 b. Honestamente, João deveria servir o bolo.
 c. Realmente, João deveria servir o bolo.
 d. Particularmente, João deveria servir o bolo.
 f. Absolutamente, João deveria servir o bolo.

Se posicionarmos mais à esquerda os elementos clivados, obteríamos a permanência dos Atos de fala na sentença. Se posicionarmos a clivagem após os Atos de fala, não alteraria a classe desses advérbios.

- (35) a. (Quem é que,) Francamente, (Quem é que) deveria servir o bolo?
 b. (Quem é que,) Honestamente, (Quem é que) deveria servir o bolo?
 c. (Quem é que,) Realmente, (Quem é que,) deveria servir o bolo?
 d. (Quem é que,) Particularmente, (Quem é que) deveria servir o bolo?
 f. (Quem é que,) Absolutamente, (Quem é que) deveria servir o bolo?

De acordo com Sant'ana (2010, p. 118-119), a natureza do núcleo que projeta os Atos de fala ocorrem diante de verbos no presente, “os advérbios desse grupo só podem aparecer

14. 14 Sant'ana (2010) classifica o “Então” como um advérbio de tempo em posição marcada, configurando uma focalização contrastiva.

com predicados em que o verbo esteja no presente quando o processo verbal é subjetivo, como ‘odiar’ e ‘amar’, sobre o qual não pode haver quantificação ou certeza absoluta”, como nos mostram os exemplos abaixo:

- (36) a. Francamente, João odeia a Maria.
b. Realmente, João ama a Maria.

Os advérbios associados ao ato de fala podem ser interpretados como orientados para o sujeito na presença de verbos no pretérito ou de ação objetiva. Com isso, perdem a força ilocucional, aparecendo em posições marcadas pré-verbais, pós-verbais e finais, demonstrando a característica dos Atos de fala em projetar ao seu núcleo a força ilocucionária (SANT’ANA, 2010).

- (37) a. Francamente João serviu o café.
b. João (francamente) serviu (francamente) o café (francamente).

Em posições que marcam a atitude ilocucional do advérbio, é mantida a característica da Modalidade Ato de fala:

- (38) a. Acredito, particularmente, em Deus.
b. Não acredito, particularmente, em Deus.
c. Acredito em anjos, particularmente.
d. Não acredito em anjos, particularmente.

Gonzaga (1997) observa que o advérbio verdadeiramente, diferente de francamente, não marca a força ilocucionária por questões restritivas mais aparentes, mesmo iniciando sentença, pois aparenta reforçar o verbo de (38) agindo como o advérbio mesmo, imprimindo a propriedade de verdade sobre o verbo, e conseqüentemente funcionando como modo.

- 39) a. ??/*Verdadeiramente, o João resolveu o problema.
b. ??/*O João verdadeiramente resolveu o problema.
c. ??/*O João, verdadeiramente, resolveu o problema.
d. O João resolveu verdadeiramente o problema.
e. O João resolveu o problema verdadeiramente.
f. ??/*O João resolveu o problema, verdadeiramente.

Sobre o exemplo acima, verificaremos na hierarquia as posições que atribuem tais restrições que posicionam os advérbios no PB. Os estudos de Cinque aplicados ao PB e analisados por Sant'ana (2010) revelam que muitos tipos de advérbios seguem o seu postulado, como os advérbios Atos de fala que não são antecipados por outros, porém, alguns não se encaixam por se fixarem em posições opostas, como vemos adiante.

(40) **Ato de Fala > Avaliativo:**

??Francamente, (in)felizmente João abriu a porta.

(41) **Ato de Fala > Evidencial:**

Realmente, dificilmente João abrirá a porta.

(42) **Ato de Fala > Epistêmico:**

Honestamente, provavelmente João abrirá a porta.

(43) **Ato de Fala > Tempo:**

??Realmente, então João abrirá a porta.

(44) **Ato de Fala > Alético necessidade:**

Francamente, necessariamente João abrirá a porta.

Classificação do Ato de Fala: (in)felizmente, *lamentavelmente*, *inevitavelmente*, *surpreendentemente*, *francamente*, *honestamente*, *sinceramente*, *acreditadamente*, *desacreditadamente*, *verdadeiramente*, *probabilisticamente*, etc.

- Posição inicial com marca prosódica e sentenças com verbo subjetivo (verbos que explicita emoções, sentimentos) no tempo presente:

(45) Francamente, João ama Maria.

- Posição pré-verbal focalizados com verbo subjetivo:

(46) João, surpreendentemente, ama Maria.

- Entre o verbo auxiliar e o principal focalizados:

(47) João vai, francamente, amar a Maria.

• Posição final focalizados:

(48) João ama Maria, francamente.

Sant'ana (2010, p.141) classifica os advérbios *particularmente* e *absolutamente* como Atos de fala (II), devido a eles se posicionarem na segunda camada funcional e ocorrerem em “sentenças com verbos assertivos e com sujeito na enunciação igual ao sujeito do processo verbal”; quando em posição pós-verbal pausado e focalizado e; em posição final, pausado e focalizado.

- (49) a. Particularmente acredito em anjos/ Absolutamente não quero ficar velha.
 b. Eu acredito, particularmente, em anjos. (íbidem, p. 141)
 c. Eu acredito em anjos, particularmente.

Advérbios Orientados para o Sujeito é uma noção originalmente classificada por Jackendoff (1972), na qual são conceituados os advérbios *francamente*, *honestamente*, *felizmente* etc, por ser mais altos, já antecipados pelo autor, quando iniciam na sentença sem marcação prosódica indicando pausa com verbo de expressão objetiva, e quando em posição pré-verbal, pós-verbal e final e, focalizado.

- (50) a. Francamente João expôs sua opinião / Honestamente João devolveu o dinheiro/
 Particularmente João conversou com a aluna / Felizmente João andava pelas ruas.
 b. João francamente expôs sua opinião.
 c. João expôs, francamente, sua opinião.
 d. João expôs sua opinião, francamente.

Nos casos em que há possibilidade de inversão, o advérbio Ato de Fala perde essa característica. Sant'ana afirma que os evidenciais e epistêmicos seguem os avaliativos, e quando estão em lados opostos, ganham uma leitura orientado ao sujeito, assim como os Atos de fala.

(51) Avaliativo > evidencial:

- a. Felizmente, obviamente João servirá o bolo.
 b. Obviamente, felizmente João servirá o bolo.

(52) Avaliativo > epistêmico:

- a. Felizmente, supostamente João comerá o bolo.

b. Supostamente, felizmente João comerá o bolo.

Os advérbios Avaliativos classificados são *(in)felizmente*, *inevitavelmente*, *(in)fortunadamente*, *lamentavelmente*, *surpreendentemente*, *curiosamente*, *inesperadamente*, etc. Ocorrem em diversas posições (inicial, pré-verbal, pós-verbal e final).

(53) a. ((in)Felizmente) (lamentavelmente) (inevitavelmente) (surpreendentemente) João abriu a porta.

b. João ((in)Felizmente) abriu ((in)Felizmente) a porta ((in)Felizmente).

(54) **Evidencial > epistêmico:**

a. ?*Evidentemente, supostamente João comeu o bolo.

b. ?Evidentemente, certamente João comeu o bolo.

O problema em (54) centra-se mais nas questões semânticas, em que o Evidencial se caracteriza por evidenciar uma experiência visual de um estado, evento ou processo descritivo na asserção, já o Epistêmico revela uma suposição possível de uma informação a ser confirmada. Ambas podem levar a valores iguais na proposição.

Na classificação dos advérbios Evidenciais *supostamente*, *aparentemente*, *obviamente*, *claramente*, *evidentemente*, *presumivelmente*, *possivelmente*, *necessariamente* etc., ocorrem em posição inicial, pré-verbal, pós-verbal e final.

(55) a. (Supostamente)(Presumivelmente)(Provavelmente)(Certamente) (Possivelmente) João comeu o bolo.

b. João (Supostamente) comeu (Supostamente) o bolo (Supostamente).

No caso do (55), segue a ordem estabelecida caso não haja oposição semântica. Nos casos em que há um alético, a ordem geralmente é respeitada, diferenciando quando há a presença de tempo, focalizando os aléticos.

(56) **Evidencial > alético:**

a. Obviamente, necessariamente João comerá o bolo.

b. *Necessariamente, obviamente João comerá o bolo.

(57) **Epistêmica > alético de necessidade:**

- a. Supostamente, necessariamente João comerá o bolo.
- b. *Necessariamente, supostamente João comerá o bolo.

Não é nítida a diferença resultante da inversão da ordem, já que o Alético estabelece verdades na proposição e o Epistêmico a certeza sobre ela baseada na informação na sentença, porém, mantemos a afirmativa de não produtividade observada por Sant'ana (2010).

Os Epistêmicos *provavelmente, supostamente/presumivelmente certamente, possivelmente, necessariamente* etc, se localizam na posição inicial e em posição pré-verbal, pós-verbal e final, focalizados.

Os advérbios de Tempo do tipo T(anterior) e T(futuro), respectivamente, classificam os itens *ontem, anteriormente* etc., e *amanhã, futuramente* etc. Além dos advérbios *agora e então* em posição inicial ou focalizados quando posicionados em pré-verbal, pós-verbal e final. Sant'ana argumenta que alguns advérbios altos, como os de Tempo, no PB, qualificam outros advérbios baixos, contrariando, aparentemente, a universalidade da hierarquia de Cinque (1999). O ordenamento oposto ocorre naturalmente, mas sem qualificação entre eles nessa inversão.

(58) Tempo (anterior) e Asp terminativo:

- a. João já não come mais carne.

(59) Tempo (anterior) e Asp retroativo/aproximativo:

- a. João já imediatamente subiu as escadas.
- b. *João já recentemente subiu as escadas.

(60) Tempo anterior e genérico/progressivo:

- a. O carro já caracteristicamente corria quilômetros por hora.

(61) Continuativo e Asp retrospectivo/aproximativo:

- a. João ainda recentemente comeu carne.
- b. *João ainda imediatamente comeu carne.

A questão, aqui observada, é a semântica que interfere na gramaticalidade da sentença em (61b). Por marcar a duração de um evento iniciado anteriormente ao ato de fala, o

Continuativo “ainda” não poderia anteceder o Retrospectivo, por este se relacionar a um evento ocorrido um pouco antes da referência de tempo.

(62) **Continuativo e Asp durativo:**

- a. João ainda longamente abre a porta.

(63) **Perfeito/imperfeito e retrospectivo/aproximativo:**

- a. João sempre imediatamente sobe as escadas.
b. *João sempre recentemente sobe as escadas.

Como dito antes, o Retrospectivo se relaciona a um evento ocorrido um pouco antes da referência de tempo, caso seja invertido, para Sant’ana (2010), não ocorre agramaticalidade, como visto em (63 a). Outro valor para “imediatamente” está ligado a questão semântica, algo como “rapidez”. É preciso observar uma ordem semântica em que a produção dessa sentença (63b) é agramatical, por ter uma característica temporal entre “sempre” e “recentemente”.

O alético de necessidade (des)necessariamente é listado por Sant’ana, mas há o modo alético de possibilidade evidenciado por Santos, que classifica o item possivelmente que não é incluído na adaptação da hierarquia funcional de Sant’ana acima. Ambos os advérbios ocorrem em posição inicial e em posição focalizado inicial, pré-verbal e final.

(64) a. (Des)Necessariamente João abrirá a porta.

- b. (Necessariamente) João (Necessariamente) abrirá (Necessariamente) a porta (Necessariamente).

2.4.3 Advérbios Baixos

A ordem dos advérbios baixos segue a hierarquia de Cinque (1999, 2006). Em relação aos advérbios Ato de fala, Avaliativos, Evidenciais, Epistêmicos e Aléticos de necessidade se fixam na posição anterior aos dos advérbios baixos, tais advérbios seguem estritamente a hierarquia, ao sofrerem uma inversão de posições que causam agramaticalidade, de acordo com Sant’ana (2010).

(65) **Habituais > acelerativos (I):**

- a. Habitualmente rapidamente João abre a porta.

b. ??/*Rapidamente habitualmente João abre a porta.

Os Aspectuais Habituais *usualmente, habitualmente, costumeiramente, regularmente, geralmente* etc, ocorrem na posição pré-verbal e focalizados em posição inicial, pós-verbal e final.

(66) a. João (habitualmente) (usualmente) (regularmente) (geralmente) come carne.

b. (habitualmente) João come (habitualmente) carne (habitualmente).

(67) **Repetitivo (I) > terminativo:**

a. João novamente não usa mais perfume.

b. ??/*João não usa mais perfume novamente.

Os Repetitivos (I) *repetidamente, novamente* etc, ocorrem em posição pré-verbal e focalizados em posição inicial e pós-verbal.

(68) a. João (repetidamente) (novamente) come carne.

b. (repetidamente) João come (repetidamente) carne.

(69) **Acelerativo (I) > terminativo:**

a. João rapidamente não abriu mais a boca

b. ??/*João não mais abriu a boca rapidamente.

Os Acelerativos (I) *velozmente, depressa, rapidamente, abruptamente, bruscamente, ligeiramente* etc, ocorrem em posição pré-verbal e focalizado em posição inicial, pós-verbal e final.

(70) a. João rapidamente comeu a carne.

b. (rapidamente) João come (rapidamente) a carne (rapidamente).

O Frequentativo (I) *frequentemente*, ocorre em posição pré-verbal e focalizado em posição inicial, pós-verbal e final.

(71) a. João frequentemente come carne.

b. (frequentemente) João come (frequentemente) carne (frequentemente).

(72) **Tempo anterior > Asp terminativo:**

- a. João já não come mais carne.
- b. *João não come mais carne já.

O advérbio de Tempo (anterior) *já*, ocorre em posição inicial e focalizado na posição pós-verbal e final.

- (73) a. João já comeu a carne.
- b. João comeu? (já) a carne (já).

(74) **Habitual > Aproximativo:**

- a. João habitualmente imediatamente sobe as escadas.
- b. ??/*João imediatamente habitualmente sobe as escadas.

(75) **Habitual > Durativo:**

- a. João habitualmente abre longamente a boca.
- b. *João longamente abre habitualmente a boca.

(76) **Retrospectivo/aproximativo > Asp durativo:**

- a. João recentemente brevemente acenou para o público.
- b. *João brevemente recentemente acenou para o público.

(77) **Habitual > Genérico/progressivo:**

- a. Carros habitualmente tipicamente correm 160k/h.
- b. ??/*Carros tipicamente habitualmente correm 160k/h.

(78) **Habitual > Acelerativo (II):**

- a. João habitualmente come a carne rapidamente.

(79) **Repetitivo > Acelerativo (II):**

- a. João novamente come a carne rapidamente.

(80) **Continuativo > Acelerativo (II):**

- a. João ainda come o bolo rapidamente.

O Continuativo *ainda*, ocorre em posição pré-verbal e focalizado contrastivamente na posição inicial, pós-verbal e final.

- (81) a. João ainda come carne.
b. (*ainda)João come(ainda)carne(ainda).

(82) **Tempo (anterior) > Acelerativo (II):**

- a. João já comeu a carne rapidamente.

(83) **Perfeito/imperfeito > Acelerativo (II):**

- a. João sempre come o bolo rapidamente.

O advérbio Perfeito/imperfeito sempre se comporta do mesmo modo que o Continuativo.

- (84) a. João sempre come carne.
b. (sempre) João come (sempre) carne (sempre).

(85) **Retrospectivo > Acelerativo (II):**

- a. João recentemente comeu o bolo rapidamente.

(86) **Genérico/progressivo > Acelerativo (II):**

- a. Carros caracteristicamente alcançam 160k/h rapidamente.

Os advérbios Genéricos/progressivos caracteristicamente, inerentemente, tipicamente etc., ocorrem em posição pós-verbal e focalizado com leitura contrastiva (ou não) em posição inicial, pré-verbal e final.

- (87) a. O carro corre (tipicamente) (inerentemente) (caracteristicamente) 160km/h.
b. (caracteristicamente) O carro (caracteristicamente) corre 160km/h (caracteristicamente).

De acordo com Sant'ana (2010), o Completivo (II) completamente segue todos os advérbios baixos (o Habitual, Repetitivo, Acelerativo, Tempo (anterior), Continuativo, Perfeito/Imperfeito, Retrospectivo, Durativo e Genérico/progressivo). O Terminativo é uma expressão que antecede a forma verbal e o segue posteriormente.

- (88) a. João não come mais carne.
b. João já não come mais carne.

Os Durativos brevemente, longamente, demoradamente, instantaneamente etc. ocorrem em posição pós-verbal e focalizados na posição inicial, pré-verbal e final.

- (89) a. João conversou (brevemente) (longamente) (demoradamente) (instantaneamente) com o filho.
b. (brevemente) João (brevemente) conversou com o filho (brevemente).

Os advérbios Repetitivo (II) repetidamente, frequentemente e novamente, ocorrem em posição final com verbo no passado e focalizado com verbo também no passado em posição inicial, pré-verbal e pós-verbal.

- (90) a. João leu o texto repetidamente.
b. (repetidamente) João (repetidamente) leu (repetidamente) o texto.

Já os advérbios Completivos completamente, totalmente etc., ocorrem em posição final e focalizados com leitura contrastiva em posição inicial, pré-verbal e pós-verbal.

- (91) a. João comeu a carne (completamente) (totalmente).
b. (completamente) João (completamente) comeu (completamente) a carne.

Apesar de não englobar todos os advérbios e a hierarquia de núcleos funcionais, as evidências listadas por Sant'ana (2010) demonstram o carácter misto da posição dos advérbios e uma possível classificação de acordo com a natureza dessa categoria, entre os tipos Modalizadores, Modo, Tempo e Aspectuais, além das não realizadas no PB segundo o autor. Sant'ana (2010) demonstra a complexidade dessa categoria lexical e funcional, apesar dos avanços sobre essa categoria.

2.5 CONSIDERAÇÕES

A importância de retomar os estudos sobre o advérbio a partir da evolução histórica e descritiva, partindo desde os gregos até as pesquisas linguísticas de cunho gerativo, se justifica por demonstrar as diversas visões sobre essa categoria, assim, deixamos claro que herdamos um problema que recai sobre a formalização linguístico computacional em relação aos diversos itens abarcados nesta categoria de comportamentos sintáticos e semânticos diversos, caracterizando uma classe bastante heterogênea.

Observamos que o legado grego constitui as bases terminológicas gramaticais e estabeleceram uma influência nos estudos da gramática tradicional. Baseando-se em Neves (2005), que utiliza fontes diretas e estudos precedentes das quais, destacamos Dionísio o Trácio, por este ter organizado de fato a primeira gramática ocidental. Vimos que a sua gramática abordava apenas aspectos da fonética e da morfologia de tradição da escola de Aristarco, que listava oito categorias gramáticas (nome, verbo, conjunção, particípio, artigo, pronome, preposição e advérbio) e classificava as palavras por critérios formais (flexão e posição). O advérbio era classificado como parte do discurso, sem traços de flexão alguma. Em comparação ao Verbo, podemos observar como eram identificados esses contrastes, a partir da tabela baseada em Neves (2005, p. 253) fica evidente as características:

	Indicações Gramaticais	
	Verbo	Advérbio
Tempo	Indica tempo	Sem Flexão
Caso	Sem Caso	Sem Flexão
Pessoa	Indica pessoa	Sem Flexão
Número	Indica Número	Sem Flexão
Voz	Praticada ou recebida	Sem Flexão
Distribuição		Antes e depois do Verbo

A formalidade que identificava o advérbio em Dionísio o Trácio, se caracterizava por a falta de flexões formais e a sua posição relativa ao verbo, tanto pré-verbal quanto pós-verbal na frase. Diferente do verbo, que possuía traços específicos e definido diretamente na palavra, e não no discurso, como o advérbio que se caracterizava como parte e não se estabelecia como uma palavra fixa, mas como um tipo de circunstancial. Como vimos anteriormente, Apolônio Díscolo, um gramático alexandrino que continuando os estudos de Dionísio o Trácio, foi o

primeiro a tratar de sintaxe e definiu o advérbio como uma palavra indeclinável, de acordo com o conceito demonstrado por Neves (2005, p. 144), o advérbio “se predica de maneira geral ou particular aos modos dos verbos e que, sem eles, não pode completar o pensamento”, indicando que o advérbio só tem sentido se estiver diante de um verbo ou particípio, diferente do verbo que possui significado completo e independe do advérbio para significar. É importante indicar que Apolônio Díscolo identificava a natureza do advérbio como um tipo de palavra formada tanto da junção das palavras (o -mente em português na construção de advérbios) quanto em formas isoladas (como a palavra “assim” ou “vamos!”), mas nessa classificação, as interjeições são postas no mesmo grupo do advérbio. Outros estudos que avançam sobre a gramática e o estudo do advérbio mantêm muito dessas características definidas pelos Helênicos, se diferenciando na classificação de tipos (modo, tempo etc.), além da herança das categorias deixadas pelos gregos, pois das dez que temos atualmente, sete já pertenciam a gramática grega antiga (nome, verbo, conjunção, artigo, pronome, preposição e advérbio). Apesar de todo o tempo que se percorreu em relação aos estudos gramaticais e a expansão do conceito de modificação do advérbio sobre outras categorias, as modernas gramáticas tradicionais mais atuais aceitaram algumas análises descritivas e alguns avanços das pesquisas linguísticas em que os advérbios ganham mais atenção diante de uma heterogeneidade complexa observada desde o início nos estudos sobre a linguagem.

Ao evidenciar as perspectivas da linguística sobre a categoria advérbio, verificamos que Jackendoff (1972) e Laenzlinger (1998) são motivados por uma análise sintática que determina a categoria a partir das características semânticas. Desse modo, a classificação dos advérbios, essencialmente, são os advérbios de frase e advérbios de VP para Jackendoff, e estes advérbios podem tomar duas projeções funcionais na sentença, de acordo com Laenzlinger. O movimento do verbo para esse tipo de análise se vale do comportamento do advérbio na sentença, pois o mesmo determina se há ou não motivação para o movimento do verbo em línguas como o inglês e o francês, por exemplo, visto em Pollock (1989) este tipo de análise. Ao procurar evidências a partir de outras categorias para elucidar o comportamento do advérbio, Pollock (1989) parte da análise detalhada do movimento dos verbos para analisar o comportamento sintático dos advérbios, pois os mesmos acompanham a categoria verbal. Em Cinque (1999, 2004, 2006) ao propor uma classificação geral para os advérbios e projeções funcionais, expandindo as projeções consideravelmente, Cinque (1999) compara diversas posições dos advérbios diante de verbos e quantificadores em diversas línguas, postulando algumas posições na geração dos advérbios.

Uma das diferenças que marcam os trabalhos desses autores gerativos são as tendências de análise dos advérbios, segundo Sant'ana (2010), há dois tipos, a primeira é a baseada no escopo que tem como teóricos Ernst (2004) e Costa (1996), na qual adota a perspectiva de que o advérbio é inserido na sintaxe sob a operação de adjunção. Já a segunda é baseada no traço, em que os teóricos Laenzliger (1998) e Cinque (1999) são os pesquisadores que adotam tal perspectiva teórica, determinando os itens lexicais com função adverbial inseridos na sintaxe por projeções passíveis de checagem de traços nas posições sintagmáticas. No nosso caso, adotamos a visão que concebe essas relações de traços.

Apesar de Jackendoff postular diversas posições para o advérbio, Cinque (1999) elabora uma classificação mais detalhada e isso colaborou para pensarmos em utilizar a sua classificação na formalização e modelação, tendo em vista que no PB, o advérbio possui uma maior mobilidade em comparação a outras línguas, como o inglês, por exemplo. O trabalho de Sant'ana (2015) parte do trabalho de Cinque (1999, 2006) nas suas pesquisas, analisando a partir de teste de avaliação gramatical e agramatical de sentenças com advérbios ditos Altos e Baixos na hierarquia cartográfica, demonstrando as possibilidades e limitações das postulações de Cinque ao ser aplicado no PB. Por estar pré-classificado os advérbios postulados por Cinque, denotam uma formalização aparentemente mais tranquila por já identificar estabelecer as posições desses itens, porém, isso não esgota as análises sobre o advérbio, por não ser um tipo de análise unânime, pois postula os advérbios projetados para ocupar a posição de especificador e não adjunto. Por isso, a modelação não deverá ter um caráter geral.

Na elaboração de um analisador sintático, é preciso ter dados no dicionário e regras gramaticais específicas 16 na sua constituição. Por isso a urgência desse capítulo, pois demonstra que há diversas possibilidades de universalizar os conceitos de modo geral sobre essa categoria tão heterogênea, apesar de ter características em comum que lhes identifiquem o pertencimento a classe. Diante das inúmeras análises linguísticas sobre o advérbio, a representação, aqui, se limitará por questões de espaço e tempo, deixando para a futura Tese o fechamento da modelação linguístico computacional e computacional dos advérbios em um parser. No momento, os dados apresentados demonstraram como são distribuídos estes itens, apesar de a representação se limitar, por vezes, ao nível dos constituintes que formam as sentenças, não significará que não são úteis para a representação dos sintagmas funcionais, pelo contrário, por buscarmos caminhos para demonstrar os advérbios hierarquizados tanto isolados quanto encaixados em sentenças simples com os traços unificados, precisamos utilizar as informações obtidas para as representações, como veremos na seção pertinente.

3 O LINGÜÍSTICO E O LINGÜÍSTICO COMPUTACIONAL

Jurafsky e Martin (2010) afirmam que as metodologias e os temas entre a linguagem e a tecnologia recebem níveis de descrição e modos de representações necessárias na construção dos recursos computacionais, sendo eles, as representações linguísticas que modelam o conhecimento linguístico com o objetivo de desenvolver sistemas de Processamento de Línguas Naturais (PLN) a serem aplicadas nos softwares destinados à manipulação da linguagem para diversos fins. Os métodos utilizados para a construção desses sistemas necessitam de informações sintáticas e semânticas sobre o léxico. Demonstramos no capítulo anterior como são organizados os itens lexicais da categoria advérbio, visto por alguns pesquisadores, predominantemente, por seus aspectos semânticos que determinam a sua localização na sentença, e por outros pesquisadores que determinam a sua posição graças a projeções funcionais. Seguimos com um conjunto de propriedades dos advérbios Altos e Baixos para descrever esta classe de palavras, e com isso, englobar a descrição e a representação formal para servir de modelo tratável por máquinas. Para isso, fornecer uma fundamentação teórica e metodologia para a descrição e formalização que a codificação necessita é o objetivo deste capítulo.

A necessidade de formalizar conceitos teóricos já eram vistos em Chomsky (1957), ao fornecer definições formais a observação dos fenômenos linguísticos, em que testes para a busca de inconsistências e insuficiência teórica eram passíveis de melhorias sobre o fenômeno em questão. De acordo com Veenstra (1998 apud MODESTO, 2011, p. 148), nas teorias com descrições informais, é mais difícil encontrar inconsistências ou problemas na aplicação ou na própria teoria utilizada na explicação do objeto pesquisado, a autora adverte que “é praticamente impossível determinar as consequências de teorias informais e, assim, de mostrá-las falsas”. Diferente das teorias formais que levam a descobertas de soluções e detectam incoerências, pois quando se utiliza linguagem formal, somos obrigatoriamente orientados a precisão, assim, “todas as noções usadas em uma definição devem ser definidas elas mesmas” (MODESTO, 2011, p. 148). Diante de uma formalização linguístico computacional de uma teoria informal, segundo Modesto, pode-se acarretar grandes erros. Apesar disso, Veenstra se mostra a favor de uma formalização gradual com linguagem que tenha especificação própria, para assim, ser traduzida em um programa de computador.

Essa busca da representação formal da linguagem humana (re)produzida por uma máquina (ou computador) é um dos objetivos da Linguística Computacional (LC), não

necessariamente seu principal motivador atual. Segundo McEnery (1992 apud SARDINHA, 2005, P. 23), a LC “é um campo de estudo que se desenvolveu a partir de um cruzamento entre a linguística, enquanto disciplina geral, e o subconjunto das ciências da computação chamado de “inteligência artificial”. E por essa intersecção, de acordo com Sardinha (2005), a criação de modelos computacionais de aspecto da língua é predominante, como atesta a ACL (Association for Computation Linguistics)¹⁵:

Linguistas computacionais interessam-se em fornecer modelos computacionais de vários tipos de fenômenos linguísticos. Tais modelos podem ser “baseados em conhecimento” (“feitos à mão”) ou “guiados por dados” (“estatísticos ou empíricos”). O trabalho em linguística computacional é motivado por uma perspectiva científica, de tal modo que se tenta fornecer uma explicação computacional para um fenômeno linguístico ou psicolinguístico; e em alguns casos a motivação pode ser mais puramente tecnológico, de tal modo que se procura fornecer um componente operante de um sistema, de fala ou de linguagem natural. De fato, o trabalho do linguista computacional está atualmente incorporado em vários sistemas operantes, incluindo sistemas de reconhecimento de fala, sintetizadores de fala, sistemas de resposta de voz, motores de busca da web, editores de texto, materiais de ensino de língua, entre outros. (SARDINHA, 2005, P. 24)

É perfeitamente plausível, na LC, um trabalho linguístico de teoria gerativa ser inserido nas pesquisas de cunho mais computacional, pois as testagens de teorias e trabalhos mais formais são passíveis de procedimentos testáveis. Isso é evidenciado por McEnery (1992 apud SARDINHA, 2005, P. 24), pois como linguista, “o pesquisador pode se preocupar em ver a linguística como aplicação de teorias linguísticas em máquinas, a fim de provar teorias em um domínio formal [...] Dentro de tal paradigma, a linguística computacional seria a ciência de interpretar a teoria linguística dentro de um domínio computacional”. A perspectiva de usar a linguagem humana para criação de programas são mais evidentes e poderemos observar que são até em maior número de trabalhos realizados.

É perfeitamente plausível, na LC, um trabalho linguístico de teoria gerativa ser inserido nas pesquisas de cunho mais computacional, pois as testagens de teorias e trabalhos mais formais são passíveis de procedimentos testáveis. Isso é evidenciado por McEnery (1992 apud SARDINHA, 2005, P. 24), pois como linguista, “o pesquisador pode se preocupar em ver a linguística como aplicação de teorias linguísticas em máquinas, a fim de provar teorias em um domínio formal [...] Dentro de tal paradigma, a linguística computacional seria a ciência de interpretar a teoria linguística dentro de um domínio computacional”. A perspectiva de usar a linguagem humana para criação de programas são mais evidentes e poderemos observar que são até em maior número de trabalhos realizados.

15. <https://www.aclweb.org/portal/>

Esses trabalhos específicos se inserem na LC por meio do Processamento Natural da Linguagem (PNL), uma das (sub)áreas ao lado da Linguística de Corpus que formam a LC. Othero e Menuzzi (2005) caracterizam essa divisão que, por vezes não é nítida, já que existem inúmeros trabalhos envolvendo essas áreas, a partir dos objetos que cada uma tem por principal. A Linguística de Corpus se ocupa dos trabalhos e pesquisas com corporas eletrônicos, utilizando banco de dados com amostras de linguagem natural, podendo ser de diferentes fontes, de língua falada (por exemplo, fala de crianças durante a aquisição, entrevistas etc.) e de língua escrita (exemplo de obras literárias, jornais, revistas, produção textual, dentre outros). Já o PNL se ocupa das construções de software, aplicativos, sistemas computacionais específicos, como tradutores automáticos, *chatterbots*, *parsers*¹⁶, reconhedores de voz, geradores automáticos de resumos, buscadores de plágios, etc.

A partir da literatura, Alencar (2006) lista os significados atribuídos a LC como (i) disciplina da linguística (assim como a sociolinguística, psicolinguística, neurolinguística etc.) preocupada com o desenvolvimento de algoritmos das línguas naturais e do PLN, o desenvolvimento de formalismos gramaticais é um exemplo; (ii) voltado para fornecer ferramentas computacionais para pesquisas linguísticas e para o processamento de dados linguísticos, no qual a LC é uma das aplicações desenvolvidas; (iii) disciplina que aborda a implementação dos fenômenos da linguagem no computador, área mais ligada ao PNL, pois se relaciona às questões da inteligência artificial e da ciência da cognição de forma mais geral; (iv) disciplina destinada à ciência aplicada, tem o objetivo de desenvolver os aplicativos de tradução, correção ortográfica e gramatical etc. Por vezes, destinado à engenharia de software, engenharia da gramática e tecnologia da linguagem (em inglês *grammar engineering ou language technology*).

Apesar dessa interação, os distanciamentos entre a linguística e as ciências da computação proporcionaram, por muito tempo, atritos por questões plausíveis entre os informatas e os linguistas. Dias da Silva (1996, p. 17) cita algumas razões para o afastamento mútuo devido à “fragmentação, a parcialidade e a pouca formalização dos modelos linguísticos são também apontadas como agravantes para o quadro de distanciamento”. Mykowiecka (1991 apud DIAS DA SILVA, 1996, p. 18) ecoa acriticamente aos linguistas ao afirmar que:

um dos motivos que vem impedindo o rápido desenvolvimento do estudo do PLN é o fato de que a maioria dos linguistas não está disposta a cooperar [...].

16. Em inglês uma análise sintática computacional é chamada de *parsing* e o mecanismo utilizado é chamado de *parser*, como não utilizamos autômatos, mas apenas gramáticas e algoritmos (linguísticos), utilizamos apenas o nome *parser*.

o que, conseqüentemente, acarreta uma escassez de teorias lingüísticas e de definições suficientemente precisas para o uso computacional (MYKOWIECKA, 1991, p. 497).

Por outro lado, o argumento contrário mais difundido entre os linguistas é a constatação de que os engenheiros da linguagem desenvolvem sistemas de PLN rudimentares e sem qualquer fundamentação lingüística. O distanciamento provocado pela PLN sobre os linguistas é justificado por Moreno Fernández (1990 apud DIAS DA SILVA, 1996, p. 18-19), ao apontar que:

esses grupos só existem para alimentar a indústria de informática: cada grupo trabalha para oferecer ao mercado consumidor programas mais sofisticados, mais eficientes e mais econômicos que os programas desenvolvidos por seus pares. Acrescenta que, por esse motivo, é muito difícil encontrar publicações que informem com regularidade e transparência os avanços alcançados por parte das instituições, sejam elas públicas ou privadas. Em outras palavras, a informação não é divulgada para o público interessado, porque isso significaria ceder resultados para um competidor potencial (MORENO-FERNÁNDEZ, 1990).

Apesar das inúmeras razões para o distanciamento, Allen (1995) afirma que os estudos lingüísticos são fundamentais para o desenvolvimento do PLN. Existem, segundo o pesquisador, duas motivações para o desenvolvimento de modelos computacionais, a científica e a consolidação da área de pesquisa. A motivação científica, que busca uma melhor compreensão de como a linguagem funciona, reconhecendo que as abordagens tradicionais não possuem ferramentas que resolvam o problema da compressão e da produção da linguagem, mesmo que se combinassem todas as teorias, seria um trabalho complexo, por isso, a utilização de computadores e programas poderiam, observando e superando as falhas de execução, melhorar gradativamente a sua aplicação. Os modelos computacionais podem fornecer previsões específicas sobre o comportamento humano, podendo ser explorado até pelos psicolinguistas, já que, com o processo, poderíamos compreender com profundidade o processamento da linguagem humana. Tal sonho só é possível realizar mediante esforços combinados da lingüística, da psicolingüística, da filosofia, e dos cientistas da computação. Este objetivo comum motiva o surgimento de uma nova área de pesquisas interdisciplinares, por vezes chamado de ciência cognitiva.

Essa preocupação sobre a interação é vista por Dias da Silva (1996) de forma sensível, pois afirma que enquanto os: lingüistas lutam para introduzir recursos da informática e das Ciências da Computação em suas pesquisas lingüísticas, projetistas de PLN lutam para domar as línguas naturais, sem poder contar com o auxílio de fundamentação lingüística adequada. Essa duplicação desnecessária de esforços não só dificulta a descoberta de soluções, que seguramente seria agilizada com o incentivo de trabalho solidário, como é também bastante reveladora, pois deixa transparecer, e até mesmo justificar, a autonomia que se instala entre os Estudos do PLN e os Estudos da Linguagem. A desvinculação das duas áreas é, porém,

extremamente preocupante, porque contribui para aumentar ainda mais os desencontros e, principalmente, minimizar a importância do papel dos lingüistas na proposição e no desenvolvimento de projetos de PLN, que, em geral, resultam de iniciativas tomadas por não-lingüistas e tornam-se privilégio de instituições ou departamentos que investem na pesquisa tecnológica. O desencontro entre pesquisadores é causa de muitas das inadequações detectadas nos estudos sobre o PLN e, sobretudo, é desestímulo para iniciativas voltadas para a criação de tecnologias lingüísticas (DIAS DA SILVA, 1996, p. 41)

Para alcançar os objetivos, Dias da Silva (1996) recomenda que os estudos voltados ao PNL devam ser divididos em três domínios: o lingüístico, o lingüístico computacional e o computacional. No nosso caso, objetivaremos os dois primeiros níveis. Com essa equação o fenômeno lingüístico a ser estudado será explicitado, analisado e a partir dos resultados da descrição, passará pela formalização lingüístico computacional, para em seguida, servir de implementação para programas específicos. Esses três domínios colaboram com a afirmação de Jurafsky e Martin (2010)¹⁷, ao esclarecer que o papel da LC na lingüística é esclarecer fatos e fenômenos da linguagem e formalizá-los para o uso em PNL.

A importância de ter uma sequência linear nos estudos que tratam dos fenômenos e aplicação por meio de computadores, em especial, para os trabalhos dos pesquisadores da linguagem natural (principalmente lingüistas), é proposto por Dias da Silva (1996, p. 89), como podemos observar, temos:

Domínios	Problemas	Recursos
Lingüístico	Conhecimento e uso lingüístico	Teorias da competência e do desempenho
Lingüístico Computacional	Representação e formalização	Linguagens formais de representação
Computacional	Codificar as representações e formalismos	Linguagens de programação e sistemas computacionais

Esses domínios listados acima são fases cujo método são conceituados por Dias da Silva (1996) a partir dos seus objetivos. Baseado nessas fases, temos os seguintes conceitos:

17. 17 Dan Jurafsky e Chris Manning fornecem curso de processamento natural da linguagem na plataforma de vídeos do Youtube, no canal OpenCourseonline, também disponibilizam os materiais vistos nas suas vídeoaulas (slides e pdfs) nos seguintes endereços, respectivamente: <<https://www.youtube.com/playlist?list=PL6397E4B26D00A269>> e <<https://web.stanford.edu/~jurafsky/slp>>

Fase Linguística	
Elaboração do conjunto de conhecimentos sobre a própria linguagem, decompondo e entendendo os fenômenos linguísticos primordiais para o desenvolvimento do sistema.	Descrição informal
	Uma sentença declarativa simples é formada por um Sintagma Verbal concatenado a um Sintagma Nominal.
	Formalismo linguístico
	$S \rightarrow NP_{\text{sujeito}} V_{\text{predicado}}$

Tabela 1: Fase Linguística (DIAS DA SILVA, 1996, p. 93)

Fase Linguístico computacional	
Empreendimento conceitual do sistema a partir da representação dos resultados da fase anterior, onde se projetam as representações linguísticas e extralinguísticas nos sistemas formais.	Traços Específicos Gramaticais
	<NP pessoa> = <VP pessoa>
	<NP número> = <VP número>
	<NP caso> = nominativo <VP forma verbal> = finita
	Regra Sintática
	$S \rightarrow NP VP$

Tabela 2: Fase Representacional (DIAS DA SILVA, 1996, p. 93)

Fase Computacional¹⁸	
“além de transformar as representações da fase anterior em programas computacionais, estudam-se as questões referentes à integração conceitual e física dos vários componentes envolvidos, questões referentes ao ambiente computacional em que o sistema será desenvolvido e implementado”	Implementação em Python/NLTK
	S[Tipo = [Declarativo]] → NP[Pessoa=[1/2/3], número = [singular/plural], caso = [2]] VP[pessoa =[1], número= [singular/plural], Caso=[nominativo], forma =[infinitiva]]

Tabela 3: Fase Implementacional (com traços formais) (DIAS DA SILVA, 1996, p. 94)

Apesar dessa linearidade ser respeitada nesse trabalho, tentaremos referenciar mais as questões linguísticas, pois o fenômeno aqui pesquisado ainda carece de melhores descrições e explicações. Ao ser formalizado na fase computacional, buscaremos não apenas explicitar as hipóteses teóricas, mas testá-las e analisá-las.

As questões linguísticas referentes aos fenômenos apresentados inicialmente no capítulo destinado aos advérbios mostraram a complexidade em obter tanto os traços sintáticos e semânticos, quanto a natureza dos advérbios desde os estudos iniciais com os gregos até a teoria gerativa, que possui discordâncias entre os estudos que admitem o advérbio como especificador de um núcleo funcional e os que concebem-no como adjuntos. Antes de formalizar as questões relativas ao advérbio, revisitaremos a teoria linguística de cunho gerativo em seu estado atual, pois ela será utilizada como guia na formalização, pois tal programa de investigação científica se apresenta desde o início (independente das suas variações e modelos) como um norteador de inúmeras pesquisas em PLN. A vantagem em utilizar os pressupostos teóricos gerativos está na adequação e robustez da formalização. O conjunto de diretrizes para a formalização que foi mencionada acima consiste, de acordo com Veenstra (1998), de uma divisão da formalização em três partes: o dicionário, uma assinatura (regras) e um axioma de traços (gramaticais e/ou formais).

18. O exemplo simplificado do uso do Python/NLTK (Linguagem de programação e sua biblioteca – *Natural Language Toolkit*) para a implementação foi feito no domínio computacional, pois no capítulo destinado ao *parser* detalharemos mais sobre essa linguagem, além dessas marcações seguirem modelos linguísticos computacionais reais.

3.1 O PROGRAMA MINIMALISTA

A linguística Gerativa é uma teoria geral, com pesquisas nos âmbitos fonológicos, morfológicos e semânticos, porém, para os teóricos gerativos, a Sintaxe ocupa um lugar central, pois o caráter gerativo das línguas é revelado no componente sintático da gramática. Gramática aqui entendida como o conhecimento inato que um falante tem da sua própria língua. Esse conhecimento possui ferramentas que formam sentenças gramaticais e não lhe permite realizar sentenças agramaticais graças a sua Competência linguística. É por meio da Competência que podemos realizar sentenças no uso efetivo da língua na interpretação e comunicação com o meio (MIOTO; SILVA; LOPES, 2013).

Para Chomsky (1995), o estágio inicial de uma criança em fase de aquisição da linguagem, a GU, é constituída pelos Princípios (conjunto de regularidades gramaticais inatas universais) e Parâmetros (variedades gramaticais moldadas de acordo com as peculiaridades de uma determinada língua). A GU, como estágio inicial, retira da língua do ambiente (Língua-E) as informações necessárias para a formatação da gramática e ao concluir, uma Língua-I(nterna) terá emergida na mente desse indivíduo, com todas as suas características estáveis, havendo mudanças apenas na aquisição de novos vocabulários.

O interesse então para a maioria dos gerativistas é desvendar como se realiza esses processos na mente de um indivíduo que habilita a formação de infinitas frases. Essas possibilidades estão condicionadas a regras, e tais regras impossibilitam a criação de sintagmas e sentenças, devendo obedecer aos critérios estabelecidos e fixados na Língua-I. Os estudos linguísticos de cunho gerativo incidem nas pesquisas de diferentes línguas, na busca de regras universais comuns em todas as línguas existentes e as regras específicas de cada língua para descobrir o que, de fato, é a faculdade da linguagem humana e de como atua na mente de um indivíduo, além de como efetua a aquisição de primeira e segunda língua.

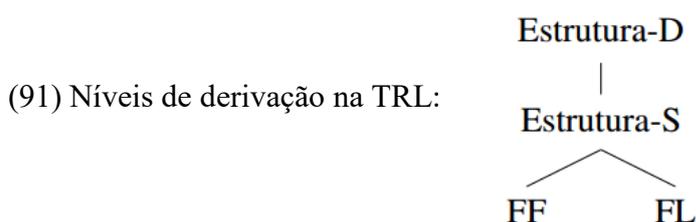
Para descrever essas operações que estruturam a língua natural, o gerativismo assume a independência da sintaxe das outras interfaces, como a semântica, pois a estrutura independe do significado. Fórmula dois tipos de metodologias, na qual a primeira são os julgamentos de aceitabilidade, e a segunda é a formalização adequada das descrições generalizáveis proposta pelo pesquisador. Em seguida, deve esse pesquisador expor o parâmetro, a regra computacional, o fenômeno sintático etc., utilizando-se de modelos teóricos sustentados por uma maquinaria (Teoria Padrão; Teoria da Ligação e Regência (TRL); Princípios e Parâmetros) de análises para a investigação dos fenômenos descobertos. Esses modelos são

métodos formais, possibilitando a descrição dos conjuntos computacionais, conhecidos como derivação, responsáveis por gerar frases, de uma maneira mais objetiva e científica (KENEDY, 2015).

A teoria gerativa evoluiu desde o seu início em 1957 com a publicação do livro *Syntactic structures* que revolucionou a área da linguística ao propor uma investigação mentalista ou mais precisamente, com Chomsky 1965 e 1986 com as ideias basilares da teoria até chegar na fase do modelo dos Princípios e Parâmetros. Entretanto, o programa de investigação minimalista (CHOMSKY, 1995) surge, influenciado por diversas pesquisas, por exemplo, as pesquisas de Pollock (1988), galgando formalismos mais econômicos e simples que os modelos teóricos anteriores, e as interfaces fonológica e lógica recebem mais atenção que antes. Aqui estaremos mais focados na fase atual do minimalismo (CHOMSKY, 2001), objetivando ao menos uma estrutura de análise do *parser*.

De acordo com Guimarães (2015), o PM é um programa de pesquisa inserido no P&P com o objetivo de enxugar todo o maquinário da gerativa, tornando-a mais explicativa e o mais simples possível sobre o funcionamento da linguagem, elaborando um conjunto de diretrizes na elaboração de modelos formais. Eliminando (ou reformulando) o que seria desnecessário, por exemplo, os níveis que uma sentença possuía, na TRL, uma combinação de itens lexicais inseridos na estrutura sintática de acordo com a Teoria X-barras e o Critério Temático, formando uma Estrutura-D que ao sofrer regras de transformações, são ordenadas de acordo com a ordem de palavras pronunciadas obtidas através de diversas operações, como as regras de movimentos, gerando a Estrutura-S (superficial).

Em P&P são definidas as relações gramaticais e a aplicação de filtros (como o filtro de Caso) que começam a geração de sentenças com estruturas agramaticais. Esses dois casos são os níveis internos da gramática, os outros dois últimos níveis são gerados a partir de mais transformações partindo da Estrutura-S bifurcando em Forma-Lógica (FL) e Forma-Fonética (FF), respectivamente, significando uma representação (abstrata) do significado da sentença, mantendo relações com o sistema cognitivo do uso e interpretação e; uma representação (abstrata) dos sons e suas relações com o sistema psicomotor da articulação e percepção.



No primeiro momento do PM, o que era assegurado na Estrutura-D (as relações temáticas e as estruturas de constituintes da X-barr) agora são garantidos pelo princípio de economia, pela numeração e a operação concatenar (*merge*, em inglês). Essa redução acompanha o pensamento de que os resultados da derivação sintática resultam na FF e FL, por esses servirem como *input* para diversos sistemas mentais (som e significado) não se poderia eliminá-los, como nos indica Modesto (2011). Agora a gramática não possui mais as representações dos níveis internos, apenas os níveis que se relacionam com sistemas extragramaticais som e significado. Representada na imagem em (92):

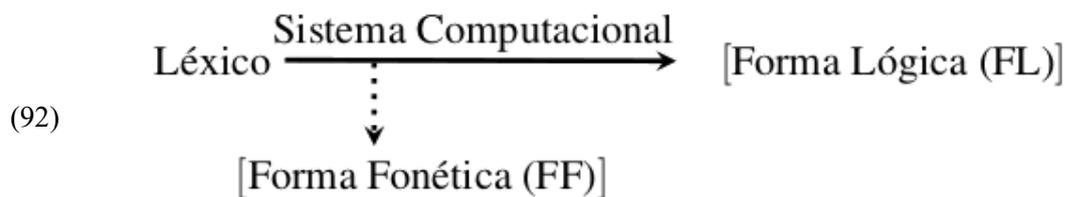


Figura 1: Arquitetura adaptada de Guimarães (2015, p. 29)

Neste esquema, as representações mentais construídas pela linguagem são derivadas por um processo computacional que opera na construção das representações, a partir das informações presentes no léxico, que serão enviadas para as interfaces. Observaremos melhor esse funcionamento mental da linguagem detalhando mais abaixo com o esquema adaptado da arquitetura da linguagem presente em Kenedy (2013) para melhor detalhamento.

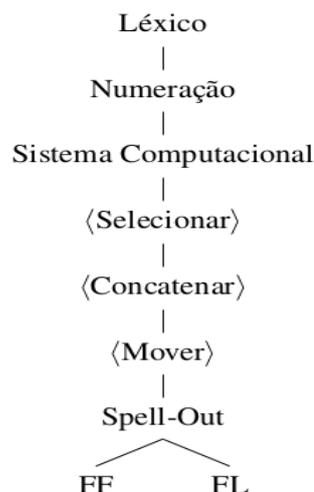
(93) Arquitetura da Linguagem com seus níveis de derivação e interfaces (KENEDY, 2015, p. 127):

curto; ou Condição do Elo Mínimo; Princípio da Procrastinação etc.). Nos limitamos apenas nas operações mais básicas no momento.

As concepções do Minimalismo advêm do senso científico que afirma que a natureza preza por sistemas mais econômicos, não se realiza por meio de caminhos sinuosos e ou difíceis. Segundo Boeckx (2006, p. 114), o “Minimalismo pode ser visto como uma investigação rigorosa da simplicidade dos princípios, realisticamente interpretado, isto é, aplicado a um objeto real, a faculdade da linguagem humana”¹⁹.

O método empregado pelo minimalismo visa menos processos (des)necessários na seleção e computação dos itens linguísticos, proporcionando uma gramática mais adequada e, essa gramática deve conter determinados princípios gramáticas que obedeçam às regras naturais preferencialmente simples de serem aplicadas, por exemplo, Numerar e Concatenar são preferenciais do que Mover durante a derivação, como afirma Hornstein, Nunes e Grohmann (2006). Após a checagem dos traços gramaticais, ao criar a estrutura sentencial no Sistema, a estrutura é submetida ao *Spell-Out*, responsável por enviar, como já referenciado, os elementos interpretáveis para a FF para receber som [π] e para a FL, recebendo significado [λ] para ser interpretado. As operações efetuadas na FF são responsáveis pelas variações e ocorrem após *Spell-Out* e as operações na FL não incidem influência marcada, já que estão indiretamente relacionados, como visto no exemplo (94) e logo abaixo, no (95), essas operações ocorrem por fases.

(94) Fases da formação da gramática no PM antes e depois de *Spell-Out*:



19. 19 No original: Minimalism can be seen as a thorough investigation of the principle of simplicity, interpreted realistically, that is, as applying to a real object, the human language faculty.

O mecanismo da gramática gerativa é formado por fases (pequenos ciclos ou partes de informações), além de incluir o sintagma verbal “vezinho” (vP), todos os papéis temáticos, a estrutura de toda a frase (CP) e suas especificações, enquanto a representação das ordens das palavras que precisam ser satisfeitas em FF e seguindo para FF na derivação são representadas no mecanismo computacional e, evidencia a preocupação do Minimalismo em ilustrar as derivações da estrutura do que a sua representação (LIMA, 2006). Algo que precisa de atenção na formulação de qualquer estruturação de um *parser* que venha a investir na utilização do PM na sua diretriz linguística.

3.2 MODELOS FORMAIS

Ao pensar em analisadores sintáticos automáticos que possibilitam representar substancialmente as análises linguísticas contemporâneas de cunho gerativo, a preocupação gira em torno de quais modelos formais utilizaremos para compor a gramática, os traços e o léxico (identificado como dicionário) neste analisador, além da linguagem de programação que irá nós auxiliar. Por isso, introduziremos alguns conceitos advindos da LC que engloba outros conceitos além do linguístico, presentes no PLN (mais difundidos na área da Computação). As análises computacionais efetuadas pelos *parsers* são orientadas pela ordem, ou seja, partindo das sequências de palavras de cada sentença à medida que são pronunciadas.

Essa característica de análise dos *parsers* não possibilita o uso dinâmico de elementos abstratos como movimento, apagamento ou cópia de forma explícitas e interativas, comuns em alguns modelos de gramáticas transformacionais. Espera-se dos analisadores computacionais, ao menos, a emulação efetiva da capacidade linguística e mental de um indivíduo, por isso, a sua gramática precisa estar de acordo com as evidências dos estudos linguísticos sobre as regras mentais da gramática (gerativa) da língua natural específica a ser utilizada no *parser*, além de um léxico produzido pelo falante com as propriedades esperadas dos itens lexicais (classe, número, gênero etc.) presentes.

Partindo das análises dos traços sintáticos e semânticos dos advérbios a partir de conceitos gerativos, a abordagem das regras presentes no PM e os modelos que se pretendem à análise de tais itens, sem deixar de lado os conceitos de Cinque (1999, 2004, 2006) sobre os advérbios altos e baixos, caracterizam as unidades presentes no dicionário. Como os advérbios são heterogêneos e, por isto, as análises diversificam (PERINI, 2005; OLIVEIRA, 1996), a unificação dos traços sintáticos e semânticos utilizando as Matrizes de valor-atributo como recurso para captar alguns traços neste item nos parece promissor devido a característica dinâmica das informações impressas nos itens por meio das MAVs. É preciso

lembrar que o movimento dos itens no PM ocorre através dos traços formais que os itens devem checar nas interfaces.

As regras gramaticais seguem as estruturas delineadas inicialmente pela Gramática Livre de Contexto (CFG, do inglês Context-Free Grammar) (OTHERO; MENUZZI, 2005; OTHERO; MARTINS, 2011; ALENCAR, 2011) enriquecida com traços e o X-barra como modelo de descrição interna sintagmática, contudo, tais modelos não seguem o modelo utilizado no Minimalismo proposto por Chomsky (1995), Chomsky (2001), a Bare Phrase Structure (em português, estrutura sintagmática nua ou estrutura sintática por fase), no máximo, seguem a antissimetria de Kayne (1994) com algumas adaptações. Na fase Computacional, a criação do parser fica a cargo da biblioteca NLTK (Natural Language ToolKit) (BIRD; KLEIN; LOPER, 2014) presente na linguagem de programação Python(3).

Como estamos na fase linguístico-computacional ou representacional, segundo Dias da Silva (1996), apresentaremos os modelos formais compostos por uma língua formal e uma gramática que guiam os tipos de algoritmos, mas que não possuem relação direta com os modelos teóricos linguísticos das línguas naturais. Uma língua formal (L) e uma gramática (G) respectivamente, no âmbito linguístico-computacional, segundo Alencar (2011), é um conjunto sequencial resultante da “concatenação de elementos extraídos de um conjunto de símbolos, chamado alfabeto, ou sigma” representada por Σ (ibidem, p. 19) e a gramática é “o mecanismo capaz de produzir todas e somente as cadeias pertencentes a essa língua” (ibidem p. 22).

Partindo do sigma $\Sigma = \{a\}$ poderemos formar a língua $L = \{a, aa, aaa\}$, graças à possibilidade de repetição e concatenação dos elementos de $L = \{a\}$ que podem ser representados por um índice nos elementos, indicando o número de repetições, $L = \{a_1, a_2, a_3\}$. Simplificando a partir de variáveis e intervalos fixos nas possibilidades que a variável pode assumir, pode-se representar como $L = a_n, 1 \leq n \leq 3$. O número de repetições raramente pode ser zero ($L = \{a_0\}$), pois uma cadeia vazia (ϵ) serve apenas como uma notação. Segundo Jeuring e Swierstra (2001), esse alfabeto representa símbolos simples (a) e uma sequência de símbolos (aa), logo, uma L começa com um alfabeto de símbolos simples e termina com uma sequência (ax), essa representação determina uma língua formal finita, como as letras do alfabeto convencional (a,b,c,...z), um alfabeto binário (0,1), um conjunto de palavras reservadas das línguas de programação (if, then, else etc.), conjunto de caracteres, um conjunto de palavras da língua natural do PB (curso, pratica, exercícios, exames etc.). A formação de algumas L são exemplificadas em Alencar (2011, p. 20-21).

- $L_1 = \{\text{cama, camarão, camarada}\}$, $\Sigma = \{a,c,d,m,o,r,\tilde{a}\}$; constituída por palavras da língua portuguesa, formadas pelas cadeias de símbolos do Σ de L_1 .
- $L_2 = \{\epsilon, a, b, aa, bb, aaa, bbb, aaaa, bbbb, \dots\}$, $\Sigma = \{a, b\}$ ($L_2 = a^n b^n$, $n \geq 0$); com características infinitas por conta dos números de elementos (a e b) que podem ser repetidos, inclusive o zero (ϵ).
- $L_3 = \{\text{um jogador de futebol, aquela jogadora de futebol, os jogadores de campo, jogadores, . . .}\}$; contém expressões do português, formados pela concatenação dos elementos de um alfabeto convencional complexo; não é possível a enumeração por ser exaustiva o número de elementos possíveis, nem a especificação precisa do critério de pertinência desse grupo, não estando formalmente caracterizado, não é possível, por enquanto, saber se os elementos realmente integram esse conjunto. É possível, como falantes, identificar o que pode ser gramatical e não gramatical ou impossíveis²⁰, advindos da concatenação desses elementos.
- $L_4 = \{\text{bochechuda, bochechudas, bochechudo, bochechudos}\}$, $\epsilon = \{\text{bochech, ud, o, a, s}\}$; formada pela concatenação de elementos de um alfabeto constituído por raízes e sufixos do português.

Cada L corresponde a um tipo de G e atua num processo de inclusão a exemplo de uma gramática de L_3 que pode ser gerada por L_2 , L_1 , contudo, não pode ser gerada por uma L_4 , como dito antes, a formação é de inclusão: $L_4 \subset L_3 \subset L_2 \subset L_1$, que correspondem a G s do tipo 3, 2, 1, 0. Uma G deve representar uma língua natural no seu nível de expressão e no processo de reconhecimento. Por isso, os tipos são caracterizados de acordo com a aproximação desses níveis. Vieira e Lima (2001) definem esses tipos como: o tipo 3 são regulares e mais restritos, tornando mais fáceis o reconhecimento, porém, são insuficientes para expressar as regras de formação da língua natural; os de tipo 2 são livres de contexto (geradas por uma CFG), possuem um grau elevado na representação da complexidade de uma língua natural, mas não são suficientes para expressar níveis de dependência, como concordância verbal ou nominal, por exemplo; as do tipo 1 são sensíveis ao contexto e resolvem o problema das dependências, mas não são expressivas em relação ao reconhecimento; as do tipo 0 são sistemas de reescrita geral. São consideradas como tipo 0 as gramáticas transformacionais que de acordo com Alencar (2008), não são aceitas pelos formalismos alternativos da gramática gerativa por não gerar outros tipos de G .

20. 20 Independente das possíveis ordenações, esses elementos sempre gerarão sentenças agramaticais (*um de campo jogadora).

Como a G de uma L é um mecanismo de reprodução e organização das cadeias da $L = \{a\}$ de modo adequado para dar conta da numeração dos itens presentes, pois os elementos de uma L correspondem, por exemplo, o léxico de uma língua natural como o PB, considerada infinita. O uso de uma G é a opção mais realista e possível de formalizar os itens na construção e análise de sentenças, como indica Alencar (2011). Na Ciência da Computação, de acordo com Jurafsky e Martin (2010), os autômatos, dispositivos responsáveis por reconhecer uma $L\{a\}$, determinam se os elementos pertencentes de fato são de L , além de gerar algoritmos simulando a recepção linguística a partir de estados finitos. Por outro lado, na Gerativa e na LC, segundo Alencar (2011), tanto autômatos quanto G s são estruturas declarativas utilizadas por algorítmicos de geração, reconhecimento e análise, esses algorítmicos podem modelar aspectos da performance. Apesar de configurar dispositivos diferentes, um autômato e uma G podem ser construídas em conjunto e de forma automática graças aos algoritmos disponíveis.

Como vimos anteriormente, o Sistema-Computacional da linguagem deve formular a gramática do falante ideal (nativo de uma dada língua) em que obtém os itens lexicais e forma, a partir da numeração, as operações que servem às interfaces. Poderemos, com ressalvas, comparar tais operações com as operações de uma G . Dentre as G conhecidas, a mais utilizada é a gramática Livre de Contexto - CFG (do inglês Context-Free Grammar) vista rapidamente acima. Uma CFG pode dar conta de alguns aspectos sintáticos de uma língua natural por este ser um formalismo limitado na modelação de fenômenos próprios da predicação e da sintaxe, por exemplo, Chomsky (1965) propôs operações que afetariam o produto de uma CFG, as operações recursivas que possibilitam uma CFG formar sintagmas que geram outros sintagmas. De acordo com Jurafsky e Martin (2010), uma CFG consiste de um conjunto de regras ou produções, e essas são expressas por meio de símbolos da linguagem que podem ser agrupadas e ordenadas em conjunto, além de um léxico. Esses símbolos são divididos em duas classes, os símbolos terminais que ficam à esquerda e os símbolos não-terminais que ficam à direita e representam as generalizações. As regras de formação dos itens nesses símbolos são ordenadas da seguinte forma: $A \rightarrow AB$; em que A é um terminal, “ \rightarrow ” pode ser “consiste de” ou “gera o” e, AB são os itens do não-terminal. Note que a repetição de A e A em ambos os casos demonstra a recursividade da operação.

Segundo a formalização de Jurafsky e Martin (2010) e Alencar (2011), uma CFG de G , é formada por um quádruplo ordenado $G = (\Phi, \Sigma, R, S)$, em que:

- i Φ é um conjunto (alfabeto) de símbolos não-terminais;
- ii Σ é um conjunto (alfabeto) de símbolos terminais (com a condição $\Phi \cap \Sigma = \emptyset$);

- iii R é um conjunto de regras ou produção (ou substituição) $\langle a, \beta \rangle$ em que a é substituível por β , a partir da forma $a \rightarrow \beta$, em que a e β são elementos de Γ^* , com a condição de que $a \neq \varepsilon$ e $a \in \Sigma$, além de β ser um item (neste caso, uma string, sequência de caracteres que formam palavras ou texto) dos símbolos de um conjunto infinito de palavras ($\Gamma = \Phi \cup \Sigma$);
- iv S é um símbolo inicial ($S \in \Phi$).

Dessa forma a representação de uma CFG por um formalismo como uma PSG²¹ (*Phase Structure Grammar*, em português Gramática de Estrutura Sintagmática), é representado com adaptações da seguinte forma por Francez e Wintner (2011, p. 15):

$$(95) \quad S \rightarrow \Sigma a S \Sigma b \mid \varepsilon$$

$$\Sigma a \rightarrow a$$

$$\Sigma b \rightarrow b$$

Para efeito didático, utilizamos, com pequenas modificações, Alencar (2011, p. 26) na aplicação da definição à gramática do tipo 2 (L 3), por ser um tipo CFG:

$$(96) \quad NP \rightarrow D N$$

$$NP \rightarrow D N PP$$

$$NP \rightarrow N$$

$$PP \rightarrow P NP$$

$$P \rightarrow \{de\}$$

$$D \rightarrow \{um, os, aquela\}$$

$$N \rightarrow \{jogador, jogadora, jogadores, campo, futebol\}$$

A representação $D \rightarrow \{um, os, aquela\}$ é uma abreviatura para (98) e equivale igualmente a $D \rightarrow um \mid os \mid aquela$. Onde \mid expressa a disjunção lógica (ALENCAR, 2011), também equivale a “ou”, por exemplo, D produz ‘um’ ou ‘os’ ou ‘aquela’.

Iniciando pelo não-terminal NP, podemos obter um sintagma substituindo os símbolos sucessivamente, de acordo com as regras, até a última substituição. Por exemplo, o sintagma “um jogador de futebol” pode ser derivado da seguinte forma:

21. 21 Mais adiante conceituaremos este formalismo.

(97)

NP	(regra a)
Det N PP	(regra b)
um N PP	(regra c)
um jogador PP	(regra d)
um jogador de NP	(regra e)
um jogador de N	(regra f)
um jogador de futebol	(regra g)

Esse modelo é bastante limitado quando se pensa em modelar fenômenos reais e mais complexos, como concordância, regência, tempo etc. A CFG sozinha não dá conta de identificar qual G gera de fato uma sintaxe gramatical, não permitindo que um algoritmo determine se uma cadeia pertence ou não a língua natural alvo da análise. Apesar de existir alguns tipos de G até mais sofisticadas²², mas dada a importância desse modelo, sintaticistas “aperfeiçoaram a CFG por meio do enriquecimento das regras dessa gramática com estruturas de traços, estruturas essas sujeitas à operações matemáticas de unificação” (ALENCAR, 2012a, p. 530). No PLN, as informações lexicais precisam ser explícitas, possibilitando consistência, evitando ambiguidades, por isso, no formalismo que expande as possibilidades das CFGs, as matrizes de atributo-valor (MAV) foram amplamente utilizadas na formalização das estruturas de traços em formalismos computacionais como na FLG e na HPSG. Apresentaremos abaixo algumas descrições sobre esses dois modelos, pois eles são considerados alternativas ao modelo padrão gerativo elaborado por Chomsky (1965 e 1986), mas sem se deter com profundidade nesses modelos, pois não serão o alvo da formalização.

A LFG (Gramática Léxico-Funcional, do inglês Lexical-Functional Grammar) foi elaborada no final dos anos 1970 com a colaboração de Bresnan (um linguista) e Ronald Kaplan (um cientista da computação) que estavam insatisfeitos com o atual modelo transformacional da linguagem e procuraram elaborar uma abordagem mais “realista”, esse modelo, tipologicamente computacional, se propõem a resultados mais consistentes sobre a compreensão e aquisição da linguagem com um maquinário robusto (BRESNAN; KAPLAN, 1982). Por ser um modelo não transformacional, baseia-se em restrições, assumindo um

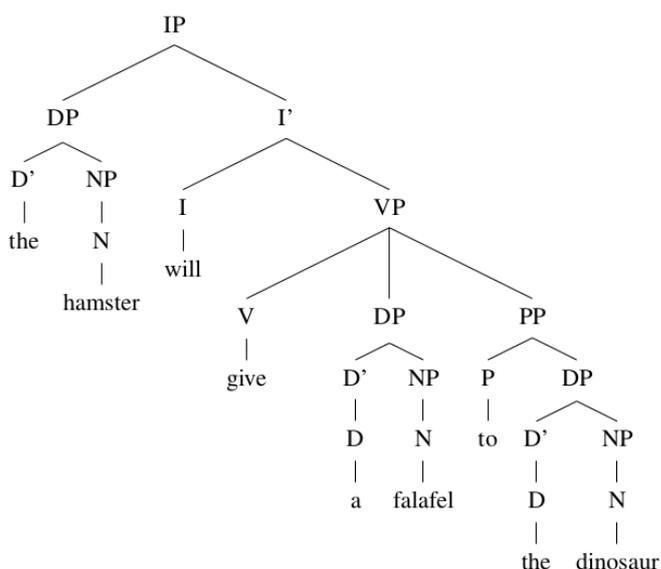
22. 22 Ver os trabalhos de Veenstra (1998), Othero e Menuzzi (2005), Alencar (2008), Dunham (2014) e Jurafsky e Martin (2010).

léxico estruturado e assume como primitivos teóricos funções gramaticais. Nesse modelo, é assumida uma abordagem relacional que estrutura os predicados e argumentos caracterizando uma análise explícita do que é universal e particular entre as línguas, com múltiplos níveis de representação organizados de acordo com as suas características.

Os principais níveis são a estrutura-A(rgumental), a estrutura-F(uncional) e a estrutura-C(onstituintes). A estrutura-A contém a informação lexical: número, tipo sintático e organização hierárquica dos argumentos de um predicador (semântica e sintática). As restrições são baseadas em princípios, como o princípio de que todo predicador possui um sujeito, por exemplo. A estrutura-F licencia as funções gramaticais como sujeito, objeto, além dos traços funcionais de tempo, aspecto, pessoa, número etc. Essa estrutura é organizada por meio de uma MAV que atribui a função e os traços a um dado item. A estrutura específica da sentença, a categoria dos constituintes, ou seja, a ordem dos itens lexicais fica a cargo da estrutura-C. Ela é representada por estrutura arbórea por meio de um tipo de X-barra, mas sem as questões de movimento ou vestígios, como em P&P. Vejamos abaixo alguns exemplos retirados de Falk (2001, p. 63), objetivando o mapeamento das construções.

(98) a. Estrutura-A: – The hamster will give a falafel to the dinosaur.

b. Estrutura-C:



c. Estrutura-F:

$$\left[\begin{array}{l} \text{SUBJ} \\ \text{TENSE} \\ \text{PRED} \\ \text{OBJ} \\ \text{OBL}_{goal} \end{array} \left[\begin{array}{l} \left[\begin{array}{l} \text{DEF} \quad + \\ \text{PRED} \quad \text{'hamster'} \text{Num} \quad \text{SG} \end{array} \right] \\ \text{FUT} \\ \text{'give'} \langle \text{SUBJ}, \text{OBJ}, \text{OBL}_{goal}, \text{OBJ} \rangle \\ \left[\begin{array}{l} \text{DEF} \quad + \\ \text{PRED} \quad \text{'falafel'} \text{NUM} \quad \text{SG} \end{array} \right] \\ \left[\begin{array}{l} \text{PCASE} \quad \text{OBL}_{goal} \\ \text{OBJ} \left[\begin{array}{l} \text{DEF} \quad + \\ \text{PRED} \quad \text{'dinosaur'} \text{NUM} \quad \text{SG} \end{array} \right] \end{array} \right] \end{array} \right]$$

Essa relação entre a estrutura-C em (99 b) e a estrutura-F em (99 c) corresponde, segundo Falk (2001) as relações dos nós correspondentes aos itens categoriais e traços formais. O valor do traço TENSE advém do I e o valor do PRED vem do V; SUBJ é uma propriedade do IP que projeta o DP, contendo os valores do SUBJ; a presença dos OBJ e OBL é graças às propriedades do VP. As operações de unificação verificam a concordância entre os traços e funções dos itens da sentença, facilitando a implementação do modelo computacional.

A HPSG de Pollard e Sag (1994) vem ganhando uma quantidade expressiva de adeptos, graças ao mecanismo descritivo robusto, capaz de explicitar, por meio de traços, as restrições e licenciamento que dado item comporta e opera sobre as demais na sentença. Tais traços carregam valores relevantes para a análise, e formaliza a sua descrição por meio das MAVs, assim como a FLG. Branco e Costa (2011, p. 219) conceituam a HPSG como um “enquadramento epistemológico para o estudo da linguagem natural humana e em particular para o estudo científico daquilo que, na sequência de uma tradição milenar, se designa por gramática”. No momento não abarcaremos as especificações desse formalismo, pois a diferença substancial entre a FLG e a HPSG são os tipos de traços e a estrutura, extremamente lexicalista não necessita das estruturas de constituintes como um componente à parte. Apesar de esses modelos serem eficientes, até sobre como poderíamos investigar os traços sintáticos pertinentes e a natureza semântica dos advérbios, graças ao poder de unificação das análises, insistimos na formalização do Minimalismo, pois segundo Rodrigues e Augusto (2009), a

LFG e a HPSG são modelos representacionais; já o Minimalismo optou por uma concepção derivacional fraca. Essa diferença tem implicações em termos da abrangência dessas propostas teóricas, fazendo com que os modelos não-derivacionais da LFG e da HPSG, de orientação lexicalista, estabeleçam uma forte aproximação com as ciências da computação. Ao assumirem um modelo matemático de funções para a caracterização das estruturas da língua, essas propostas têm se apresentado como modelos plausíveis para implementações computacionais. Já o modelo formal gerativista, o Minimalismo, dada sua natureza derivacional e a incorporação de questões de ordem conceitual na própria concepção do modelo, tem apontado para diálogos promissores com modelos de processamento. O fato de o Minimalismo ser um modelo derivacional permite que a própria caracterização da derivação sintática seja tomada, dadas as adaptações necessárias, como proximamente equivalente à formulação e ao *parsing* de enunciados linguísticos em modelos de processamento on-line (Ibidem, p. 146).

No Minimalismo, Rodrigues e Augusto (2009) afirmam que o léxico ganha cada vez mais tenção. Apesar de o Sistema Computacional ser o responsável pela formação das sentenças, as entradas léxicas em relação aos traços o aproxima da LFG e da HPSG. Apesar das diferenças entre esses modelos no aparato mecânico e formal, alguns pontos de convergência podem ser utilizados no processo de descrição e análise dos advérbios, mantendo, claro, as questões do Minimalismo como principal aporte linguístico, fundamentando, assim, a formalização do fenômeno aqui descrito.

Como vimos anteriormente, a CFG pode dar conta de algumas descrições sintáticas, mas são extremamente limitadas, Pollard e Sag (1994), ao descreverem sobre a CFG, afirmam que Chomsky propôs as regras de transformação para ampliar o poder de análise das operações sintáticas, incidindo sobre as árvores sintagmáticas, gerando novas árvores. A aplicação dessa proposta não foi utilizada, tendo em vista que a utilização dos modelos transformacionais foram preteridos pelos modelos LFG e HPSG²³. Neste cenário, a CFG apenas serviu como um esqueleto enriquecido de traços.

As propostas que seguem uma descrição nos moldes da CFG na implementação de modelos formais para implementar computacionalmente gramáticas do PB utilizam dentre dezenas de tipos de modelos formais, o formalismo DCG (Definite Clause Grammar, em português gramática de Cláusula Definida) e o formalismo PSG, já referido acima, são os mais utilizados, de acordo com Vieira e Lima (2001) e Alencar (2011).

A DCG é recurso da linguagem de programação Prolog 25 definido como um formalismo representacional aos moldes da CFG que torna mais fácil implementar gramáticas

23. 23 Não apenas por suas vantagens, mas por não terem as mesmas dificuldades de implementação das gramáticas transformacionais.

e desenvolver *parser* elaborados e criados com a linguagem de programação Prolog. A DCG foi a primeira implementação importante para a CFG na década de 80, introduzidas por Pereira e Warren (1980) e utilizada por um grande número de trabalhos linguístico computacionais, como afirma Alencar (2008). Na DCG, podemos utilizar traços por meio da unificação para dar conta de alguns fenômenos, a exemplo de concordância e regência. Esse recurso converte as regras linguísticas sintagmáticas em notações lógicas para o Prolog, com isso, ela descreve a sintaxe de uma língua natural. O uso da DCG para a descrição do PB está presente nos trabalhos de Othero e Menuzzi (2005), Othero (2006), David (2007) e David e Othero (2011).

A DCG é recurso da linguagem de programação Prolog²⁴ definido como um formalismo representacional aos moldes da CFG que torna mais fácil implementar gramáticas e desenvolver *parser* elaborados e criados com a linguagem de programação Prolog. A DCG foi a primeira implementação importante para a CFG na década de 80, introduzidas por Pereira e Warren (1980) e utilizada por um grande número de trabalhos linguístico computacionais, como afirma Alencar (2008). Na DCG, podemos utilizar traços por meio da unificação para dar conta de alguns fenômenos, a exemplo de concordância e regência. Esse recurso converte as regras linguísticas sintagmáticas em notações lógicas para o Prolog, com isso, ela descreve a sintaxe de uma língua natural. O uso da DCG para a descrição do PB está presente nos trabalhos de Othero e Menuzzi (2005), Othero (2006), David (2007) e David e Othero (2011).

A PSG é um formalismo de base gerativa, visto em Chomsky (1957, 1965), e é um tipo específico de formalismo da CFG bastante útil por gerar regras explícitas e apresentarem a estrutura sintática das sentenças de acordo com seus constituintes. Como regras básicas em (101 a) temos como resultado (101 b):

(99) a.

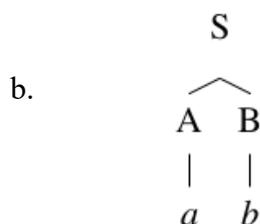
$$S \rightarrow A B^{25}$$

$$A \rightarrow a$$

$$B \rightarrow b$$

24. 24 Abreviação de Programming in Logic.

25. 25 É possível A ser igual a B ($A = B$), por exemplo, um NP pode gerar outro NP.



No exemplo (100 a), temos uma estrutura simples de uma PSG livre de contexto em que o “S” representa o símbolo inicial, o A e B representam os não-terminais e o a e b são os símbolos terminais. No exemplo (101 b), podemos observar que o S se expande em A e B; A gera a e B gera b, nessa regra de rescrita, temos uma instrução para um *parser topdown*, que inicia a construção partindo do S até o a e o b (DAVID, 2007). Esse formalismo serve como uma instância da X-barra, pois configura uma representação dos constituintes de uma sentença e a X-barra, como um módulo da gramática encarregado de mostrar a estrutura sintagmática, também demonstra a hierarquia desses constituintes. Segundo Miotto, Silva e Lopes (2013, p. 51), a x-barra “é necessária para explicitar a natureza do sintagma, as relações que se estabelecem dentro dele e o modo como os sintagmas se hierarquizam para formar a sentença”, tendo em vista que um sintagma é formado a partir de um núcleo, ela descreve a organização interna, por exemplo, do sintagma adverbial, em que seu núcleo é composto por um advérbio.

Como vimos, a PSG livre de contexto possui características gerativas e é utilizado em ambientes computacionais. Apresenta uma estrutura sintática da sentença a partir dos seus constituintes; já a DCG é um formalismo presente na linguagem de programação Prolog que se diferencia, além dos níveis de representação superiores da PSG, tem uma notação diferente. Na formalização da sentença a menina usa vestido, de acordo com esses formalismos utilizados por Pereira e Shieber (1987), Veenstra (1998) e Pagani (2004), teríamos, considerando uma estrutura canônica simples para fins de exemplificação, um sintagma nominal (NP) e um sintagma verbal (VP) e seus constituintes:

- (100) $S \rightarrow NP VP$
 $NP \rightarrow D N | N$
 $VP \rightarrow V NP$
 $D \rightarrow (a)$

$N \rightarrow (\text{menina}) (\text{vestido})$

$V \rightarrow (\text{usa})$

$(S(NP (D a) (N menina)) (VP (V usa) (NP (N vestido))))$

Convertendo para a DCG simples:

(101) $s \rightarrow NP, VP.$

$n \rightarrow d, n.$

$VP \rightarrow v, NP.$

$d \rightarrow [a]; [o].$

$n \rightarrow [\text{menina}]; [\text{vestido}].$

$v \rightarrow [\text{usa}].$

$(s(NP(d(a),n(menina)),VP(v(usa),NP(d(o),n(vestido))))).$

Para que esse formalismo em DCG dê conta dos traços que restringem a formação de sentenças agramaticais e, diferente da PSG que pode formar estruturas não pertencentes a língua, a DCG tem recursos mais aprimorados para evitar tal geração errônea. Para isso, ela deve ser modificada com a inclusão de argumentos, como gênero e número. Seguindo, em partes, uma gramática elaborada por Pagani (2004), teríamos, com algumas modificações:

(102) $s([NP,VP]) \rightarrow NP([_,Num], NP), VP(Num, VP).$

$NP(Conc, N) \rightarrow n(Conc, N).$

$NP(Conc, [D, N]) \rightarrow d(Conc, D), n(Conc, N).$

$d(Conc, D) \rightarrow [D], d(Conc, D).$

$d([\text{masc,sing}],o).$

$d([\text{fem,sing}],a).$

$d([\text{masc,plur}],os).$

$d([\text{fem,plur}],as).$

$n(Conc, N) \rightarrow ([N], n(Conc, N)).$

$n([\text{masc,sing}],menino).$

$n([\text{fem,sing}],menino).$

$n([_,sing],vestido).$

$n([_,plur],vestidos).$

$s([NP, VP]) \rightarrow NP([_, Num], NP), VP(Num, VP).$

$VP(Num, V) \rightarrow v(Num, i, V)$.

$VP(Num, [V, NP]) \rightarrow v(Num, t, V), NP(_, NP)$.

$v(Num, Val, V) \rightarrow [V], v(Num, Val, V)$.

$v(sing, t, usar)$.

$v(sing, t, usa)$.

No Prolog a notação seria “?- s([a,menina,usa,vestido],[,]).”, tendo como resultado o “yes” impresso na tela, demonstrando que a formação pertence aos argumentos estabelecidos. Caso fosse “?- s([o,menina,usar,vestidos])”, o resultado seria um “no” ao verificar que a estrutura formada não atenderia ao que foi especificado e não realizaria uma operação arbórea.

O que está expresso nesse formalismo é a formação do S por NP e VP, em que o NP possui dois valores, um qualquer (expresso pelo $_$) e um número (Num). O VP também apresenta os valores Num que serão unificados com o valor de Num do NP, tornando o S gramatical, seguindo as regras, temos uma variável Conc que representa as variáveis de concordância e número presente nos itens lexicais (sing, de singular e plur, de plural). No VP observamos uma variável “i”, marcando um verbo intransitivo, e um “t”, marcando um verbo transitivo, no caso, o verbo usar. Podem parecer simples essas marcações, mas elas têm um poder de explicitar tanto regras restritivas quanto traços morfológicos de uma língua²⁶, Pagani (2004) alerta que uma PSG, por mais sofisticada que pareça, não daria conta de todos os fenômenos estruturais linguísticos, tampouco tem condições de demonstrar certos fenômenos na sua estrutura, Pagani cita o exemplo de Shieber (1985) sobre o que ele chama de dependência cruzada, que consiste em dois constituintes diferentes com subconstituintes intercalados, por exemplo, sentenças com o advérbio “respectivamente” em “João e Maria são, respectivamente, médico e advogada”. O problema estaria, segundo Pagani, no NP ter um N (João) concordando com o N (médico) do segundo constituinte NP (do predicativo, na notação tradicional), e assim por diante, revelando a limitação do PSG por não expressar tal condição. Diferente da DCG, que graças aos seus mecanismos difere substancialmente da PSG e alcança tal condição de detalhamento formal.

Há inúmeros formalismos computacionais para análises e descrições dos fenômenos expressos na linguagem humana, tais formalismos são utilizados por diversos *parsers*, dentre

26. Para uma introdução sobre a criação de parsers utilizando a notação da DCG de forma mais complexa, ver os trabalhos aqui já referenciados, como Othero e Menuzzi (2005), Othero (2006), Pagani (2004) e o trabalho de Pereira e Shieber (1987).

eles os formalismos de base CFG estruturados por FLG, HPSG (organizados por MAVs), PSG e DCG listados aqui. Demonstramos para efeito didático essas possibilidades, porém, como dito antes, a utilização da linguagem computacional Python e a biblioteca NLTK abre possibilidades maiores de formalização, além de seguirmos o modelo arbóreo como base para formalizar a estrutura e o modelo cartográfico para a descrição dos advérbios nas sentenças. Com isso, é preciso descrever o que seria este *parser*.

3.2.1 O parser, a gramática e o léxico

Nesta seção apresentaremos alguns conceitos em relação ao parser, o tratamento do léxico e a estruturação gramatical. Iniciaremos com conceitos relevantes, partindo de concepções linguísticas que dão base a qualquer processo linguístico computacional. Tomando como base Branco e Costa (2011), a faculdade da linguagem é uma capacidade cognitiva e autônoma em relação a outras capacidades cognitivas. Os estudos gramaticais objetivam descobrir o conhecimento linguístico implícito a essa capacidade cognitiva. É por meio das teorias gramaticais que o conhecimento é explicitado de forma segura e precisa.

Diante de uma gramática que representa o conhecimento que o falante dispõe acerca do número ilimitado do conhecimento linguístico das entidades pertencentes à língua natural utilizada por esse falante competente, a representação dessa gramática é uma teoria compatível com um algoritmo de parser. De acordo com Branco e Costa (2011, p. 221), na “teoria do conhecimento linguístico, a gramática oferece a melhor compatibilidade com a modelação do comportamento linguístico e em particular com o processamento de informação linguística”. A compatibilidade com o parser, mesmo não tão eficiente por conta da complexidade computacional, dentro das condições normais da execução do parser mental, a gramática é tratada para dar conta dos diferentes tipos de processamento linguístico, como a tradução ou produção, por exemplo.

Para uma efetiva descrição do conhecimento, Branco e Costa (2011) afirmam que a gramática precisa ter neutralidade procedimental de um dispositivo descritivo declarativo. Com isso, a compatibilidade de modelos descritivos se tornam eficientes diante dos submódulos da gramática, como a fonologia, morfologia, sintaxe, semântica e a pragmática. Uma gramática é precisa ao observar as diferentes línguas naturais, pois o núcleo comum e universal desse conhecimento é identificado, como vimos anteriormente, por Gramática

Universal, especificando regras comuns na descrição das línguas naturais e possibilitando a formalização por uma G de uma L , visto também anteriormente.

3.2.2 O Parser

O desenvolvimento de sistemas (programas e aplicativos linguísticos) são o objetivo da LC, que vem, de acordo com Othero e Martins (2011, p. 100), “se dedicando à descrição, à formalização e à emulação computacional das habilidades linguísticas humanas”. Ainda de acordo com esses autores, a LC toma o comportamento linguístico passivo de mecanização, com isso a análise sintática, do ponto de vista da LC, é um objetivo pertinente e relevante para os trabalhos de desenvolvimento de sistemas computacionais que tratam da linguagem natural.

Dentre os diversos *parsers* já desenvolvidos para a língua portuguesa, temos o Palavras, o Curupira, o LxParser e, o Grammar Play. O Palavras²⁷ (BICK, 2000) é o parser de maior abrangência e a principal ferramenta de análise sintática automática do PB. Já o Curupira (MARTINS et al, 2003) é o parser utilizado no interior do ReGra²⁸ (Revisor gramatical) automático do PB desenvolvido pelo Núcleo Interinstitucional de Linguística Computacional – NILC, da USP de São Carlos. O Grammar Play (OTHERO, 2006) foi desenvolvido com o objetivo de ser um aplicativo didático e de fácil utilização, podendo ser utilizado em cursos de introdução à sintaxe formal, LC e PLN. O LxParser²⁹ é o serviço online de demonstração de um *parser*, é descrito no próprio site como um apoiador no ensino, na investigação e no desenvolvimento do domínio da ciência e tecnologia da linguagem natural.

Diante de um modelo formal de gramática (G), é preciso um formalismo que articule o conhecimento lexical e as regras gramaticais de estruturação dos itens da língua. O *parser*, ao analisar uma sentença, determina se tal sentença é gerada, a partir de um processo algorítmico, por essa gramática. Caso seja afirmativo, atribui a ela uma estrutura baseada nas regras formalizadas. A palavra *parsing* vem da palavra latina *pars orationes*, significando partes do discurso, “e se refere aos procedimentos mentais que determinam a estrutura de uma frase, parte integrante dos processos de produção e compreensão da linguagem” (MAIA; FINGER, 2005, p. 15). A definição na LC diz respeito ao processo de análise sintática

27. 27 Mais informações podem ser encontradas na página do Visual Interactive Syntax Learning: <http://visl.sdu.dk/>.

28. 28 O ReGra é o revisor gramatical presente desde 2000 no pacote de aplicativos Microsoft Office.

29. 29 Disponível em: <http://lxcenter.di.fc.ul.pt/services/en/LXServicesParser.html>

(semi)automática de sentenças das línguas naturais por meio dos *parsers* e possuem uma grande importância nos estudos de PNL por carregar modelos ou norteadores de aplicações que visem a construção de sistemas direcionados à linguagem humana. Os *parsers* podem variar de acordo com a sua perspectiva de análise, que são, de acordo com Othero e Martins (2011), definidas pelas relações estabelecidas com o usuário, com os recursos disponíveis e a estratégia de análise.

Das relações com o usuário, as ferramentas de *parsers* variam no grau de automatismo, podendo executar uma análise sintática sem a intervenção do usuário, resolvendo a ambiguidade lexical e semântica, estruturando toda a sentença. Eventualmente, o usuário é requerido diante de ambiguidades insolúveis e estruturas mal formadas ou para melhorar a análise da ferramenta, caracterizando uma aplicação semiautomática do parser, mas isso tudo dependerá da finalidade da ferramenta em questão. Normalmente, as ferramentas de tradução auxiliadas pelo usuário não esperam a colaboração do usuário objetivando bons resultados, porém, encontramos tradutores (como o Google Tradutor) que fornecem opções para o melhoramento da tradução. Também encontramos outras ferramentas, como as de correção gramatical, que não oferecem tal opção, dispensando as possibilidades das entradas no domínio de princípios gramaticais dos usuários que nem sempre estão impressas nas formas cultas adotadas por essas ferramentas. Outra relação ligada ao usuário destacada por Othero e Martins (2011) está relacionada ao tipo de análise, probabilístico ou determinístico. Quando determinístico, o *parser* efetua como saída (*output*) uma única estrutura para a sentença analisada; enquanto probabilístico, apresentam para a mesma sentença, as possibilidades de análises possíveis, organizadas de forma hierárquica segundo a probabilidade de ocorrência.

Da relação com os recursos disponíveis, o *parser* de uma língua natural exige a disponibilidade de um léxico e de uma gramática (princípios). Além desses componentes, poderá surgir a necessidade de um conjunto de estratégias de regularizações e desambiguação lexical e sintática, como visto acima. Um dicionário a ser acessado pelo *parser* é requerido, e ele pode ter formatos variados, podendo ser formado por uma lista contendo desde morfemas, palavras, locuções, expressões e frases inteiras até uma estrutura composta por formas (sub)categorizadas, como explica Othero e Martins (2011). Ainda de acordo com esses linguistas, a complexidade do *parser* estará ligada à quantidade e qualidade das informações presentes no dicionário. Normalmente as formas presentes no dicionário são associadas as categorias lexicais como requisito mínimo para o processamento sintático automático, enquanto as gramáticas formais normalmente requerem as categorias lexicais como instâncias dos símbolos não-terminais. O léxico é obtido, normalmente, de um *corpus* anotado,

etiquetado ou apenas listado para uso de identificação da sua forma base (classe gramatical, por exemplo).

Com o *parser* de posse dos recursos disponíveis (do léxico e de uma gramática adotada), falta definir a estratégia de análise. A análise poderá ser realizada de diversas maneiras, podendo ser de forma realista, ou seja, da forma que lemos e falamos, da esquerda para a direita, mas também da direita para a esquerda. Além de baixo para cima e de cima para baixo, ou de forma combinada. Lembrando que a estratégia de análise depende, normalmente, do tipo de gramática adotada (OTHERO; MARTINS, 2011).

De forma horizontal, a estratégia mais utilizada é a que segue a linearidade da língua, no caso do PB, da esquerda para a direita, tornando a estratégia mais realista do ponto de vista fonológico e mental. Assim como na fala e na escrita, o processamento natural ocorre no instante em que estamos processando uma sentença ou frase, como já visto anteriormente ao abordarmos o *parser* linguístico humano.

Na forma vertical, também a análise ocorre, geralmente, da esquerda para a direita, mas agora a estratégia é dividida entre alguns tipos de ordem, como o decrescente e o ascendente, dos quais o *top-down* e o *bottom-up* são os mais comuns, sendo considerados as estratégias gerais para o desenvolvimento de *parsing*. O *top-down* parte do símbolo inicial para a construção da sentença, percorrendo os não-terminais até os terminais, finalizando na representação arbórea da sentença analisada. Há um certo realismo psicológico, como afirma Othero e Martins (2011), além de ser uma estratégia que possibilita a recursividade, diminuindo o conjunto de regras no *parser*. Tomando como base o exemplo de Santos (2009, p. 50), uma regra como $DP \rightarrow D N$ permite que o analisador automático gere de um DP, um D e um N. Cada uma das categorias é gerada por outras categorias usando as categorias que estão à direita da seta nas regras de D e de N, até chegar aos respectivos itens lexicais.

Porém, de acordo com Othero e Martins (2011), a recursividade traz problemas de controle, afetando o desempenho do *parser*, causando labirintos sintáticos onde a resolução se concretiza após um número excessivo de *backtrackings* (retrocessos). Outro problema do *top-down* é em relação à estrutura, pois necessitamos traçar toda a estrutura sintática da sentença, mesmo sendo prevista pela gramática adotada. Tendo esses problemas em vista, a estratégia *bottom-up* se torna uma opção frente ao *top-down*.

O *bottom-up* faz a análise de baixo para cima (forma ascendente), parte dos nós terminais (itens lexicais), passando pelos nós não-terminais até o rótulo inicial. O *bottom-up* também apresenta alguns problemas, como as regras de generalização, que possibilita a demonstração das fronteiras sintagmáticas, que segundo Othero e Martins (2011, p. 112), sem

“a visão do conjunto, a identificação das fronteiras se torna um problema nada trivial, que geralmente envolve a análise da sentença por núcleos, exigindo, pois, um formalismo de unificação”. Isso quer dizer que para implementar um parser em que a estratégia de análise é a *bottom-up*, é preciso construir as estruturas parciais, e “a partir da aplicação de regras de combinação dessas estruturas, chega-se à estrutura de toda a sentença” (OTHERO; MARTINS, 2011, p. 113).

Segundo Othero e Martins (2011), a análise *bottom-up*, normalmente, necessita de duas gramáticas, em que uma operará sobre as categorias lexicais e as categorias funcionais nucleares e a outra sobre as projeções das categorias funcionais.

Entre o *top-down* e o *bottom-up*, há o *chart* (tabela) como uma opção que combina os dois movimentos formando um *parser* híbrido mais eficiente. Essa eficiência se dá por lançar duas operações paralelas, ao lançar dois *subparsers* em direções opostas, em que as convergências são fixadas ou de modo sequencial, deixando uma estratégia de análise suspensa até que o *parser* tenha dados suficientes do movimento contrário, para, assim, tomar a decisão. A diferença, segundo Vieira e Lima (2001, p. 20), é que “o analisador tabular tem condições de lembrar as subestruturas já analisadas e, se um retrocesso for necessário, a repetição pode ser evitada”, assim a operação é mais econômica por não recomeçar a análise do zero quando uma estrutura não for realizável em qualquer lugar das etapas em curso como acontece também com a *left-corner* que combina o *top-down* e o *bottom-up* como o *chart parser*, mas realiza essa quebra de análise caso haja alguma má formação.

Com a estratégia de análise, os recursos disponíveis e a relação com o usuário definidas, o próximo passo para quem deseja analisar os fenômenos sintáticos através de um *parser* é fundamentar a teoria na fase linguístico computacional para a construção da aplicação e por meio dos recursos computacionais disponíveis, como a já mencionada linguagem Prolog que é amplamente utilizada. Mas como adiantamos, não usaremos o Prolog, mas a linguagem de programação Python e a biblioteca *Natural Language Toolkit* – NLTK. O NLTK é o pacote mais abrangente e amigável de PLN e é constituído por um conjunto de ferramentas para a análise automática de textos nos mais diferentes níveis (BIRD; KLEIN; LOPER, 2014).

3.2.3 A Gramática

Vieira e Lima (2001) afirmam que a maioria dos trabalhos em LC foram desenvolvidos de acordo com a perspectiva teórica gerativa. Tal perspectiva foi fundamental para as vertentes gramaticais citadas anteriormente (FLG, HPSG, DCG, PSG), porém, observamos que as formalizações que tentam utilizar o Programa Minimalista ainda são em menor número³⁰, a exemplo dos trabalhos encontrados que dão ênfase no PM como Santos (2009) e Veenstra (1998) que procuram evidenciar o fenômeno linguístico das suas pesquisas a partir dessa perspectiva. Apesar desses trabalhos, abarcar a perspectiva gerativa do PM sem algumas adaptações necessárias devido às limitações conhecidas das ferramentas que estruturam o parser é muito difícil, pois ainda não há um modelo que possa ser utilizado para analisar, a partir da linearidade dos constituintes, as sentenças a partir do léxico à gramática. Veenstra (1998) buscou detalhar o PM da melhor forma para abarcar as especificidades do programa, assim como Santos (2009), Veenstra prezou pela análise dos movimentos verbais, mas utilizou a DCG e a estratégia *head-corner*, identificando o núcleo e a partir dele, estruturando o movimento de constituintes na sentença. Aqui utilizaremos as mesmas ferramentas utilizadas por Santos (2009) devido a nossa proposta linguístico computacional, formalizando a estrutura gramatical por uma CFG enriquecida com traços.

Como vimos no modelo da gramática do PM em (96), são privilegiadas as operações computacionais, por não haver regras específicas a serem adquiridas. Na visão minimalista, o léxico carrega toda a informação característica de uma língua natural e o sistema codifica graças à concepção de que “os itens lexicais constituem conjuntos de traços que retratam tanto propriedades fonéticas e semânticas, como também propriedades gramaticais, representadas por traços formais do tipo: gênero, número, pessoa, Caso, QU, etc” (RODRIGUES; AUGUSTO, 2009). Com isso, é preciso definir algumas regras a partir do léxico, pois a estrutura a ser definida no *parser* está totalmente condicionada ao item lexical que deflagará toda a análise.

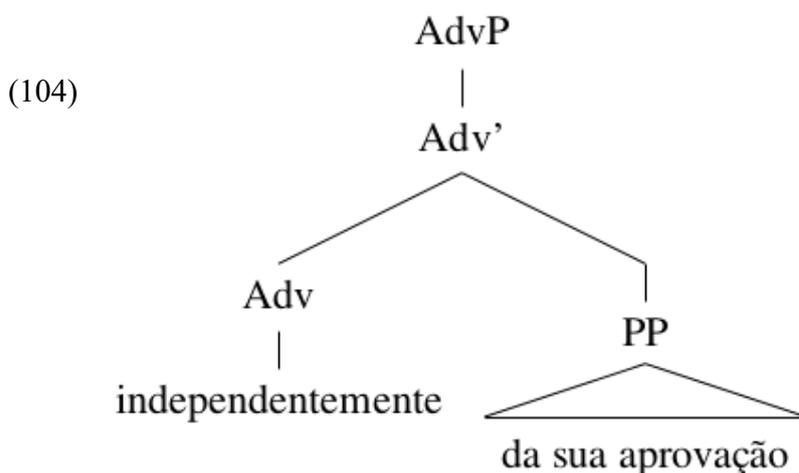
A importância da construção da gramática em seguir os pressupostos da teoria gramatical a ser utilizada e modelada, pode ser exemplificada a partir da formalização do sintagma adverbial e Othero (2006) e Othero (2009b), ao formalizar de acordo com uma gramática PSG, ele projetou o advérbio interno ao seu sintagma em (106), sem formalizar

30. 30 Ver o trabalho de Ladeira (2010) que demonstra o número de trabalhos e os tipos desenvolvidos.

qualquer tipo de possibilidade do núcleo funcional ser visível na análise. Esse movimento, por estar determinado a seguir tal configuração gramatical do parser estabelecido pelo autor, é visível a incompatibilidade com os pressupostos atuais do gerativismo. Para projetar o advérbio para a posição do seu núcleo funcional e seguir as evidências de Cinque e aplicadas por Sant'ana (2010) para modelagem na gramática de um parser que segue o PM, é preciso algumas modificações.

- (103) a. AdvP \rightarrow Adv'
 b. Adv' \rightarrow AdvP Adv'
 c. Adv' \rightarrow Adv
 d. Adv' \rightarrow Adv PP

Essas quatro regras em (103) formalizadas por Othero (2009a) que estruturam o sintagma adverbial na gramática são suficientes para a linearidade da qual o Sistema Sensorio-Motor utiliza na organização da fala. Mas nada dizem sobre o advérbio em diversas posições, sua restrição e seu papel na sentença, veja o exemplo abaixo:

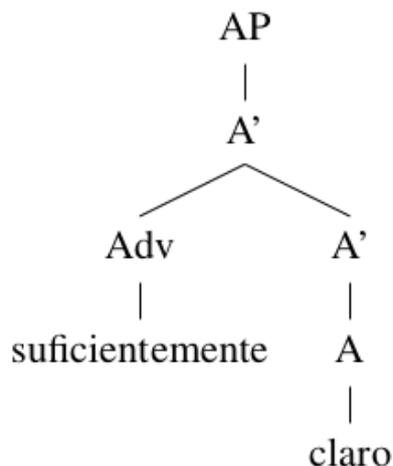


Othero (2009a) formaliza algumas regras em que o advérbio está acompanhando outros itens lexicais e seus sintagmas correspondentes, como os modificadores verbais, por exemplo, assim é ampliado consideravelmente as projeções até o nível IP. O pesquisador utilizou algumas noções da tradição linguística para estabelecer alguns posicionamentos dos advérbios na sentença como adjuntos, também estabeleceu posições fixas próximas aos núcleos, por não adentrar com mais aprofundamento em relação aos estudos sobre os advérbios, estabeleceu,

naturalmente, as relações comuns dos constituintes adverbiais. Vejamos alguns exemplos abaixo.

(105) a. Sintagma Adjetival: $A' \rightarrow Adv A' \mid A' \rightarrow A' Adv$.

b.



(106)

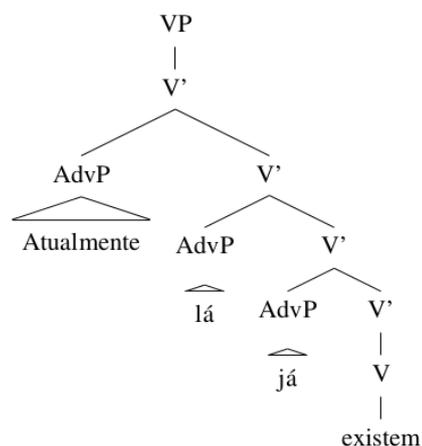
a. Sintagma Preposicional: $P' \rightarrow AdvP P' \mid P \rightarrow P' AdvP$.

b.

Sobre o Sintagma Verbal, os advérbios estão distribuídos mais livremente graças à mobilidade sintática no VP e a formalização das duas regras, nas palavras de Othero (2009a), um AdvP modifica um VP à esquerda de V' e, outra regra que prevê um AdvP modificando um VP à direita de V', possibilitando então uma ocorrência mais complexa na periferia externa e interna do VP. Vejamos abaixo a regra e a aplicação (p. 112 e 115):

(107) a. Sintagma Verbal: $V' \rightarrow AdvP V' \mid V' \rightarrow V' AdvP \mid V' \rightarrow V AdvP$.

b.

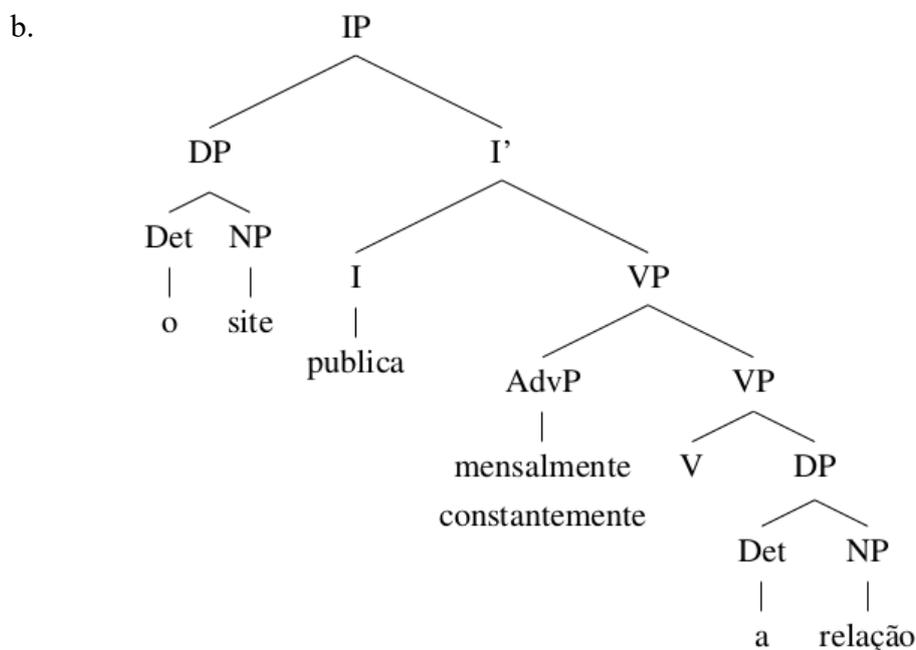


Essa análise de constituintes estrutura a ordem dos advérbios, mas sem atribuir os elementos funcionais que licenciam, tanto essa categoria quanto a dos verbais. Ainda assim, possibilita análises profundas sobre tais constituintes em relação a ordem linear, porém, para uma análise gerativa moderna, precisaria de mais elementos. No trabalho de Santos (2009), são formalizadas as posições dos advérbios e as regras para visualizar os movimentos visíveis e os não visíveis dos constituintes em sentenças finitas até a categoria funcional flexional IP. O autor modela as sentenças nas ordens categorias que o PB pode assumir, além da canônica que são SVO, SAdvO e SAdvVO, mas, por questões da limitação imposta à pesquisa, formaliza apenas as seguintes regras para abarcar os advérbios na sua análise:

- (108) a. $VP/?x^{31} \rightarrow AdvP VP/?x$.
 b. $VP \rightarrow VP$.
 c. $AdvP \rightarrow Adv$.

Sem adentrarmos com mais detalhes sobre as regras formalizadas por Santos (2009), temos a árvore (110 b) criada a partir da sentença (110 a):

- (109) a. O site publica mensalmente/constantemente a relação.



O trabalho de Santos (2009) procura, especificamente, formalizar uma minigramática para ser aplicada nas análises do parser de acordo com alguns pressupostos básicos do PM, para, assim, analisar sentenças que dê conta dos movimentos do verbo, contudo, contém as

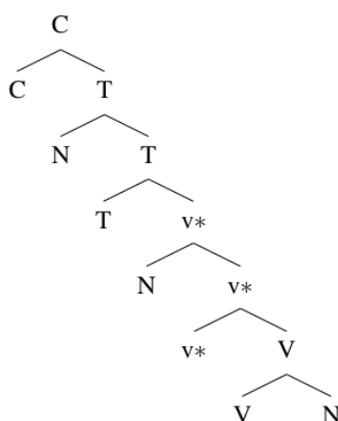
31. ³¹ O /?x simboliza um elemento ausente que será preenchido por algum traço característico de alguma categoria, no caso os traços de flexão do verbo.

mesmas limitações naturais por não ter uma formalização tão dinâmica do PM a partir dos itens lexicais.

Em estruturas que admitem adjunção à esquerda, as regras devem ser rearranjadas e incrementadas não apenas no sintagma, mas na formalização de núcleos funcionais que tornem ao menos explícito o movimento sintático. Observando uma estrutura sintática crua (sem traços unificados de interpretação do tipo EPP, Caso, gênero e número, por exemplo), poderíamos planejar a realização dos advérbios na sentença, seguindo um exemplo de estrutura canônica dos núcleos que preveem a movimentação dos verbos, para então pensarmos como definir as posições dos advérbios, teríamos:

- (110) a. $C \rightarrow CT$
 b. $T \rightarrow NT$
 c. $T \rightarrow T v^*$
 d. $v \rightarrow n v^*$
 e. $v \rightarrow v^* V$
 f. $V \rightarrow VN$

A partir dessas regras que projetam os núcleos lexicais e funcionais, poderíamos analisar uma sentença afirmativa como “João viu Maria” (KENEDY, 2013, p. 159), marcando os itens com seus traços funcionais e lexicais [C [C] [T João N [T T [v N t [v* v* [V viu V Maria N]]]]]]. Por consequência, a análise deverá realizar uma árvore do tipo (111):



(111):

A justificativa para esses traços está no fato de que traço V do item “ver” seleciona dois argumentos (interno e externo) nominais que possuem propriedades que são checadas pelo Sistema Computacional e enviadas para o Sistema de Pensamento (lógico e semântico).

As projeções (C, T, v, V) do item lexical “ver” são responsáveis pelo movimento do verbo, explicitando tal movimento na árvore sintática com os componentes básicos, abaixo temos a seguinte tabela adaptada de Fong (2008):

Item Lexical	Propriedades	Traços não-interpretáveis	Traços interpretáveis
* (transitivo)	seleciona(V) spec(seleciona(N)) valor(Caso(acc))	pes(), num(), gen() (EPP)	
V (transitivo) (inacusativo)	seleciona(N) seleciona(T)		
T	seleciona(v) valor(Caso(nom))	pes() num() gen() EPP	
C	seleciona(T)		
N (Referencial)		Caso()	pes(P) num(N) gen(G)

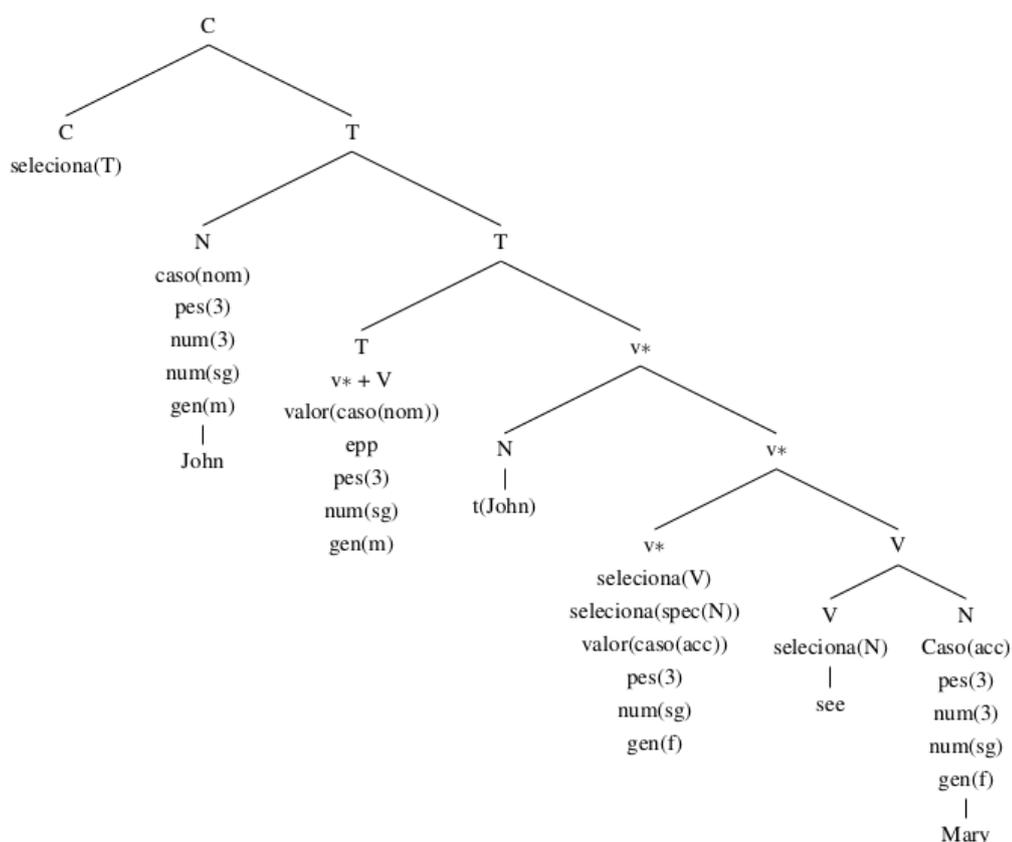
Tabela 4: Propriedades e traços interpretáveis e não interpretáveis (FONG, 2008, p. 242).

Na tabela 4, as regras de formação exigidas por um verbo transitivo demonstram as fases de movimentação sintática que esse item é submetido para checar seus traços formais (categoria, + predicador e estrutura argumental: __V__). Seguindo a explicação de Modesto (2011, p. 156), ele argumenta que é preciso tratar de forma mais detalhada os traços morfossintáticos codificados nas categorias funcionais. Como podemos observar, os traços de Caso em v e T, como também os traços de finitude (que recebem os traços de tempo e ou de concordância), exercem a função de verificador de traços, impondo restrições aos valores que devem preencher seus espaços. No caso, os valores de concordância em N externo (João) que funciona como sujeito e, do V (ver) devem ser verificados em T para conferir se seus traços são compatíveis entre si e receber o Caso nom (nominativo).

O mesmo é válido para o objeto presente no N interno (Maria), os seus valores são checados pelo V e é atribuído o valor de acc (acusativo) imposto pelo núcleo T que seleciona v* e restringe-lhe conferindo traços que, de acordo com Santos (2011, p. 156), na checagem

local, “se verificarem preenchimentos adequados dos traços, então todos os traços não interpretáveis (...) são eliminados”. Para os casos de traços interpretáveis como os traços phi (por exemplo, traços de gênero, de número e de pessoa), estão acessíveis tanto na FL quanto na interpretação semântica. Vejamos o exemplo da estrutura de Fong (2008), o verbo *see*, (“ver”, no PB), se mantém na base da árvore por não ter traços morfológicos para serem checados.

(112):



Se a verificação estiver associada ao movimento visível refletido na ordem linear dos constituintes, ou a verificação dos traços formais ocorrer após *spell-out*, devido a força dos traços formais atribuídos a forma verbal rica morfológicamente, atribuindo ao V, a adjunção e a T, exercendo influência na movimentação desse núcleo, o verbo em inglês seria adjungido, como não há traços para serem checados, o verbo permanece na posição de base. Essa fixação que define a posição do verbo infinito está, também, relacionado as categorias advérbio que lhe segue, como vimos em Pollock (1989).

Diante dessa descrição da teoria gramatical de cunho gerativa, o PM para estruturar a gramática do parser deve ser o primeiro passo. O passo seguinte será a definição do léxico, ou seja, do dicionário que comporá o parser, mas tal lista de palavras dependerá dos advérbios classificados no capítulo anterior, por enquanto não apresentaremos tal seção definindo tais

itens e seus traços, tanto para os advérbios, quanto para os itens das outras categorias. Seguiremos agora para a seção destinada à formalização e definição da linguagem Python e da biblioteca NLTK, pois será utilizada para formalizarmos o parser, tal escolha se deve, como explicado anteriormente, por conta da facilidade de aprendizagem e da abrangência dessa linguagem de programação em tocar nas questões que envolvem os estudos da língua natural no computador.

3.2.4 A linguagem Python e o NLTK

Continuando com as estratégias de pesquisas elaboradas por Dias da Silva (1996), após reunir as representações complexas do conhecimento linguístico e aplicá-las no processo de resolução de problemas linguísticos computacionais, chegamos na etapa introdutória da codificação e implementação, em que vale apresentar com mais detalhes a linguagem de programação Python e a biblioteca NLTK³². Com essas ferramentas, será possível a codificação das representações linguísticas, obtidas no nível linguístico ao explicitar o conhecimento e o uso da língua natural a partir da teoria responsável pelos estudos da competência e do desempenho. Posteriormente, apresentando o nível de representação através das linguagens formais que organizam tanto as regras quanto a pertinência do léxico.

Para desenvolver um parser é preciso uma linguagem de programação e um programa ou interpretador para essa linguagem. Apresentamos duas opções, o Prolog e o Python³³ em conjunto com a biblioteca NLTK. Optamos pelo Python, por possuir o NLTK.

O NLTK é uma biblioteca escrita na linguagem de programação Python. Essa linguagem, de acordo com Bird, Klein e Loper (2014), é uma linguagem simples, mas poderosa, com ótimas funcionalidades para o processamento de dados linguísticos, por isso a escolha do Python para implementação do NLTK. Tal linguagem possui uma curvatura de aprendizagem fácil, sua sintaxe e semântica são transparentes, além de fornecer funcionalidades na manipulação de *strings* (cadeias de palavras). Por ser uma linguagem do

32. 32 As versões do Python e do NLTK, respectivamente são 3.6.3 e 3.2.5. O ambiente de uso é o Manjaro Linux, uma distribuição baseada no Arch Linux e, os programas IDEs jupyter notebook e o IDLE3 são utilizados na implementação do *parser*.

33. 33 Borges (2014) explica que a linguagem Python “foi criada em 1990 por Guido Van Rossum, no Instituto Nacional de Pesquisa para Matemática e Ciência da Computação da Holanda (CWI) e tinha originalmente foco em usuários como físicos e engenheiros” e “foi concebido a partir de outra linguagem existente na época, chamada ABC”. Segundo Borges (2014), hoje em dia o Python é usado por diversas indústrias de tecnologia como a Google, Yahoo!, NASA, dentre outras.

tipo *scripts*, é possível uma exploração interativa também fácil e, ser orientada a objetos, os dados e métodos utilizados no Python podem ser facilmente encapsulados e reutilizados. Com uma extensa biblioteca padrão, o Python contém módulos para tarefas específicas, por exemplo, para programação gráfica, processamento numérico, processamento de dados da web, aprendizagem de máquina, criação de *bots*, assistentes virtuais etc. Por isso ele é utilizado, segundo Bird, Klein e Loper (2014), na indústria, nos estudos científicos e na educação em escala mundial.

Com a implementação do NLTK³⁴ no Python, é possível construir programas de PNL, graças as funcionalidades dos módulos para a criação de corpus e acesso; para os processos de separação e organização das sentenças e dos componentes do léxico presentes (*tokenizers*, *sentence tokenizers*); classificação e etiquetagem dos componentes da linguagem (*part-of-speech tagging*, *tag*), úteis na identificação das categorias lexicais, tanto em palavras isoladas quanto em textos mais longos, de forma automática; na criação de *parsers* dos tipos tabular, base de traços, unificação, probabilístico, dependência etc (*Chart*, *feature-based*, *unification*, *probabilistic*, *dependency*); para interpretações semânticas com cálculos lógicos, ordem lógica e checagem de modelos; criação de aplicações gráficas para concordância, *parser*, *chatbots* etc.; dentre outros. Originalmente criado em 2001, o NLTK foi desenvolvido como parte do curso de linguística computacional do Departamento de Computação e Ciência da Informação da Universidade da Pensilvaniana e, com mais desenvolvimento e apoio, se tornou uma ferramenta mais robusta, não se limitando apenas a uma ferramenta didática, mas com potencialidades de uma ferramenta de desenvolvimento (BIRD; KLEIN; LOPER, 2014).

Na análise das estruturas linguísticas, Bird, Klein e Loper (2014) afirmam que o uso do Python3-NLTK proporciona uma abordagem teórica e prática sobre os dados presentes na sentença, conferindo consistência na representação dessa estrutura formalizada. Com módulos utilizados na criação e manipulação das informações linguísticas estruturadas no nível da representação, pode-se representar o processamento de análise das expressões por meio arbóreos; estruturas de traços com seus valores unificados; utilização da CFG; e, como adiantado, o parser para receber entradas (*input*) para serem analisados automaticamente.

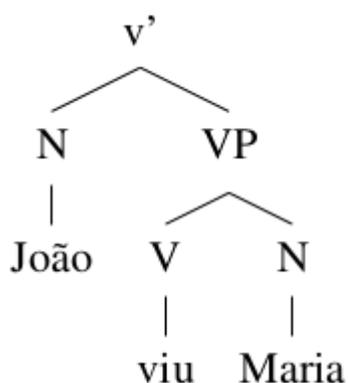
Com a estrutura e a sentença formalizadas, a partir deles a G(L) é formada, o próximo passo é repetir o que já tínhamos estruturado em (113), mas agora verificamos no aplicativo

34. 34 A instalação e utilização do Python-NLTK foram feitas em um sistema Manjaro Linux, sistema esse baseado no Arch Linux. A sua instalação neste sistema pode ser feita via terminal com o comando “sudo pacman -S python-nltk”, tal comando efetua tanto a instalação quanto a configuração para o usuário. Nos sistemas Windows, o procedimento pode ser conferido em <http://nltk.org>

estruturado com os módulos do NLTK um exemplo de uma representação utilizando a sentença “João viu maria”:

(113) a. “João viu Maria”.

b.



Nesse caso, precisamos particionar as projeções funcionais para termos uma noção melhor de como estruturar para o parser as regras estabelecidas formalmente que funcione, sem riscos de formar regras absurdas ou impróprias que venham a não realização da análise das projeções ou seja, que venha a analisar estruturas agramaticais. Seguindo Fong (2008):

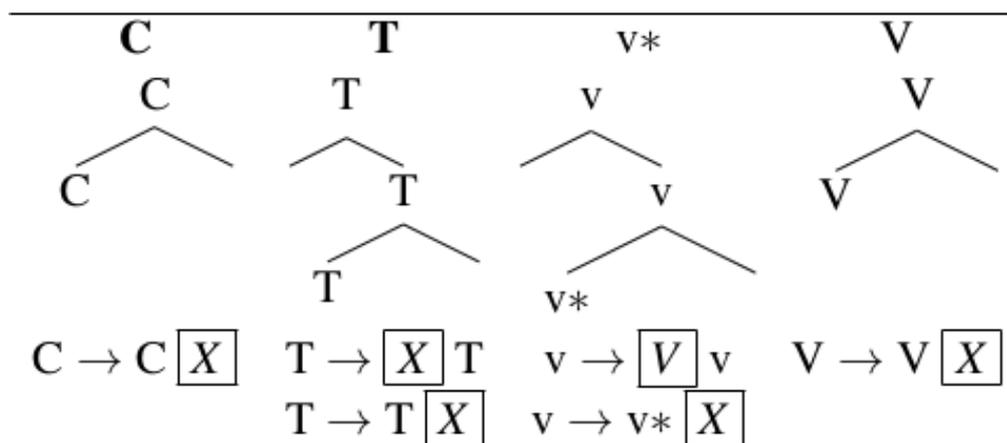


Tabela 5: Projeções funcionais para o preenchimento da G adaptadas. Fonte: Fong (2008).

Podemos observar na imagem que a exigência para criar um aplicativo pode ser mínima. Em alguns passos elaboramos um conjunto de procedimentos para representar uma CFG do tipo PSG em Python-NLTK como nos mostra Bird, Klein e Loper (2014). O comando `import` importa módulos ou objetos (classes e funções), no caso, o NLTK é

habilitado com as funções e classes gerais nele presentes, em seguida atribuímos um valor ao “g” a função CFG.fromstring que receberá uma lista de valores presentes na *multiline strings* (três aspas duplas), no caso, receberá valores sequenciais terminais e não terminais. O “sent” recebe o valor da *string* representando a sentença e os valores presentes em “g”, o *split()* é uma função para string que separa as unidades de acordo com os espaços presentes definidos. Atribuímos a variável “parser” o valor de estratégia de análise da CFG *Chart* por meio do *ChartParser* e indicando a quem receberá tal função, por fim, utilizamos o comando “for” que atribui a cada elemento (“in”) a sua atribuição corrente na variável e executa em blocos a parte determinada pelo comando anterior, atribuindo ao “tree” esses valores correspondentes. A função *print* imprime os dados atribuídos ao “tree” na tela ((SC (C) (ST (T) (VP (v) (vbar (N João) (VP (V viu) (N Maria)))))))))³⁵ e a função *draw()* desenha a possível saída de “tree”, que no caso, é uma estrutura arbórea.

A sequência de terminais e não terminais seguem uma ordem rígida, os elementos não podem ser estruturados sem conexões uns com os outros, essa dependência na modelagem das regras possibilita seguir a determinação da CFG e averiguar se o conjunto de dados presentes no dicionário e as regras formalizadas e representadas são compatíveis entre si e se estão de acordo com a teoria gramatical linguística, como podemos observar mais adiante.

Apesar de realizar análises, o parser estruturado com regras do tipo PSG, não dá conta de vários fenômenos. Com a possibilidade de gerar combinações agramaticais, o enriquecimento de traços nos itens lexicais e funcionais são necessários para evitar equívocos, segundo Santos (2009), para dar conta de fenômenos sintáticos próprios da teoria, como traços morfológicos que atribuem gênero, número e pessoa, que não são características dos advérbios. No NLTK, a função responsável pela atribuição e unificação desses traços para gerar uma árvore como a de (115) enriquecida com a estrutura de traços, é a função CFG.fromstring (FCFG), utiliza-se no lugar do formalismo gramatical clássico da CFG, que por meio do NLTK, uma estrutura de unificação FCFG, o FeatureGrammar.fromstring.

Di Felippo (2004) explica que a unificação de traços diz respeito a uma estrutura de traços que divide um dado linguístico em partes menores, os chamados atributos ou traços, para cada atributo é associado valores, por exemplo. A concordância apresenta dois atributos, número e gênero, e os valores desses atributos, respectivamente podem ser: singular (sg) e masculino (masc). São organizadas e representadas por uma MAV que explicita tais características de um dado fenômeno, estes traços são considerados restrições específicas que

35. 35 A impressão linear da árvore e os limites são postos entre parênteses () e não entre colchetes [].

integram uma gramática. Uma MAV separa os atributos e os valores por coluna, a da esquerda os atributos e os da direita os valores, como dito anteriormente, é utilizada na HPSG e na FLG, por exemplo. A MAV possibilita explicitar dados de um item lexical, ou seja, demonstrar questões morfossintáticas e, até os traços semânticos podem ser demonstrados em conjunto.

- (114) Atributo e valor – concordância (AGR); traços de número singular (NUM = sg) e; gênero masculino (GEN = masc):

$$\text{CONC} = \begin{bmatrix} \text{NUM} & < sg > \\ \text{GEN} & < masc > \end{bmatrix}$$

- (115) Código fonte simples no Python-NLTK para um DP com seus traços em MAV:

```
import nltk
from nltk import grammar, parse
g = """
% start DP
DP[AGR=?a] -> D[AGR=?a] N[AGR=?a]
D[AGR=[N='sg', P=3]] -> 'este' | 'aquele'
D[AGR=[N='pl', P=3]] -> 'estes' | 'aqueles'
D[AGR=[N='pl', P=1]] -> 'nos'
D[AGR=[P=2]] -> 'tu'
N[AGR=[N='sg', G='m']] -> 'menino'
N[AGR=[N='pl', G='m']] -> 'meninos'
N[AGR=[N='sg', G='f']] -> 'menina'
N[AGR=[N='pl', G='f']] -> 'meninas'
N[AGR=[N='sg']] -> 'estudante'
N[AGR=[N='pl']] -> 'estudantes'
"""

grammar = grammar.FeatureGrammar.fromstring(g)
tokens = input('Escreva uma Sentença: ').split()
parser = parse.FeatureEarleyChartParser(grammar)
trees = parser.parse(tokens)
print('Arboreo :')
```

```

for tree in trees: print (tree)
print ( ' ')
tree.draw()

```

Seguindo Bird, Klein e Loper (2014), no exemplo (116) é atribuído um valor ao “*grammar*” com o módulo *FeatureGrammar.fromstring* responsável por tratar da variável “*g*” que possui varias linhas graças as aspas triplas que confere uma sequência maior de valores, seguido da variável “*tokens*” que receberá a entrada do usuário graças a função *input*. O próximo passo é atribuir a variável “*parser*” o tipo de estratégia utilizada, no caso, é o ascendente (chamado de *botton-up chart parser*, pois durante o processo ele imprime na tela as regras e o léxico logo após o nome *Feature Bottom Up Predict Combine Rule*) denominado no NLTK como *FeatureEarleyChartParser*.

O resultado é a possibilidade de analisar sentenças que estejam de acordo com o fenômeno da concordância. Claro que limitado a algumas sentenças por ter um pequeno número de léxico (não-terminal) e uma única regra, o “DP -> D N”, enriquecido com traços de concordância, o *parser*, ao analisar uma sentença que não está de acordo com a regra formalizada e com isso, interpretada como agramatical, não imprime na tela a análise no modelo estabelecido, pois não encontra coerência entre os traços e o resultado não é o correto, acarretando um erro no código, do contrário é exibido a análise de acordo com a estratégia definida em conjunto da função *draw()*, responsável pela exibição arbórea no NLTK, como aponta Bird, Klein e Loper (2014). Abaixo alguns exemplos do funcionamento desse processo de análise.

(116) [Escreva uma Sentença](#): aquele menino

Linear:

```

(DP[AGR=[G='m', N='sg', P=3]]
(D[AGR=[N='sg', P=3]] aquele)
(N[AGR=[G='m', N='sg']] menino))

```

Além do resultado linear, a estrutura arbórea da sentença “aquele menino” é produzida, como mostra a imagem abaixo, com os traços determinados para cada item lexical e categoria, delineados anteriormente em (115):

(115):

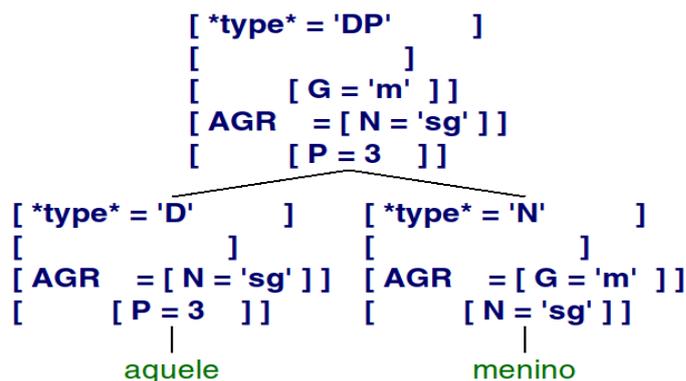


Figura 2: Estrutura arbórea elaborada no NLTK da sentença: aquele menino

Seguindo a perspectiva metodológica da gerativa, de acordo com Santos (2009, p. 78-79), é fundamental para a “pesquisa do linguista um programa como esse criar sistemas computacionais que sirvam de modelo para a descrição do conhecimento linguístico dos falantes/ouvintes”. Tal empreendimento necessita de uma elaboração mais precisa para alcançar o objetivo proposto, no caso, formalizar sentenças com advérbios de acordo com a teoria gerativa, buscando uma análise mais precisa dos mesmos, a partir dos traços que estruturam os dados com informações de qualquer natureza linguística mais formais. Com suporte do NLTK na construção de expressões e análises de sentenças a partir do *input* do usuário, seguindo a formalização da gerativa e de algumas noções advindas da Teoria Cartográfica sobre a natureza e funções dos advérbios na sentença (CINQUE, 1999; CINQUE, 2004b; CINQUE, 2006), acreditamos que seja possível uma modelagem mais significativa a respeito das sentenças com advérbios, apesar de o mesmo não possuir traços gerais, diferente das outras categorias.

O próximo capítulo se destina a evidenciar os caminhos da formalização dos itens lexicais no dicionário e as regras que deem conta da ordem do advérbio, especificaremos com traços a sua ordem hierárquica pertencente (Modalidade, Aspecto, Tempo, Modo etc.), mantendo as unidades lexicais e categoriais nas suas versões macros (CP, TP) sem implodí-las, respeitando a estrutura do PM. Demonstraremos os advérbios altos e baixos com as suas características cartográficas com traços e categorias próprias, como visto nos capítulos destinados aos estudos dos advérbios. Cobrindo então uma lacuna observada nos *parsers* sobre os advérbios na estrutura, assim, colaboramos um pouco mais nas tentativas de emular o Sistema Computacional humano, ou seja, a capacidade sintática mental nos analisadores sintáticos.

3.3 CONSIDERAÇÕES

A inserção de algumas noções fundamentais para a sistematização dos conhecimentos linguísticos de cunho gerativo são fundamentais para a construção de sistemas de PNL que desejam emular o conhecimento linguístico. Principalmente nos que se detêm na versão minimalista proposta por Chomsky (1995, 2001, adiante) para a formalização e criação de analisadores que se utilizam deste modelo que concebe o léxico, formado no sistema morfológico da língua e organizado por meio do sistema computacional mental, é estruturado. A partir dessa estrutura que se constrói as unidades sintáticas e então são enviadas as interfaces fonéticas e semânticas responsáveis pela produção e significação das sentenças. A descrição é importante devido a interação da Linguística com formalismos responsáveis por sistematizar os fenômenos linguísticos, tanto que os modelos gramaticais computacionais utilizam essas sistematizações nos sistemas de PNL, por isso, é preciso ter o rigor do método nas descrições e análises. O arcabouço gramatical é requerido por o PNL para que o desempenho e a competência sejam emulados e explicitados por meio de linguagens formais que veiculem mais informações sobre essa capacidade de produzir e entender frases/texto em diversos aplicativos.

Apesar desse capítulo abarcar tanto o linguístico computacional (representacional) e o computacional (codificação) de maneira geral, é evidente que os fenômenos linguísticos detalhados na fase linguística, e projetadas na fase representacional via modelos formais (a exemplo do PM, HPSG e FLG), exigem uma série de aplicações mais rigorosas na fase computacional. Portanto, a descrição exigida das estruturas linguísticas, como na HPSG e na FLG, utilizam mecanismos formais para descrever as entradas lexicais, além das relações dos componentes da sentença para evidenciar as restrições linguísticas e as regras morfológicas são providências para o sucesso das análises de um parser, porém, o advérbio não figura como uma categoria amplamente abarcada nas pesquisas que tratam desses modelos. O PM não é privilegiado com um estudo que tenta saturar esta categoria, e há poucos trabalhos consistentes na sua representação e codificação, até porquê a complexidade envolvida precisa de uma formalização complexa para os modelos. Aqui a formalização aplicada tenderá à unificação de traços sobre os advérbios Altos e Baixos como um modelo inicial na tentativa de elucidar mais as características desta (multi)categoria heterogênea.

Apesar desse capítulo abarcar tanto o linguístico computacional (representacional) e o computacional (codificação) de maneira geral, é evidente que os fenômenos linguísticos detalhados na fase linguística, e projetadas na fase representacional via modelos formais (a

exemplo do PM, HPSG e FLG), exigem uma série de aplicações mais rigorosas na fase computacional. Portanto, a descrição exigida das estruturas linguísticas, como na HPSG e na FLG, utilizam mecanismos formais para descrever as entradas lexicais, além das relações dos componentes da sentença para evidenciar as restrições linguísticas e as regras morfológicas são providências para o sucesso das análises de um parser, porém, o advérbio não figura como uma categoria amplamente abarcada nas pesquisas que tratam desses modelos. O PM não é privilegiado com um estudo que tenta saturar esta categoria e, há poucos trabalhos consistentes na sua representação e codificação, até porquê a complexidade envolvida precisa de uma formalização complexa para os modelos propostos. Aqui a formalização aplicada tenderá à unificação de traços sobre os advérbios Altos e Baixos como um modelo inicial na tentativa de elucidar mais as características desta (multi)categoria heterogênea.

4 ANÁLISE E MODELAÇÃO COMPUTACIONAL

Este capítulo, destinado à modelização dos advérbios, está dividido em duas seções, a primeira demonstra a representação e formalização das estruturas sintáticas modeladas de acordo com a CFG, visando uma expansão para traços unificados; a segunda seção é destinada a decodificação das sentenças visando a aplicação em um analisador sintático automático. Reiteramos a importância desta modelação tendo em vista a responsabilidade do linguista em fornecer modelos formais que representem a estrutura da língua(gem) para serem utilizadas em sistemas de PLN.

Na elaboração de um *parser* estruturado por regras de CFG e unificado por traços, é preciso reproduzir as regras gramaticais a partir das evidências linguísticas, no qual, o advérbio precisa estar inserido. As construções sintáticas que precisam ser formadas devem estar da forma mais simples possível, em contrapartida, precisam ser as mais eficientes, evitando etapas desnecessárias, mas contendo as categorias funcionais necessárias. No quesito categorias funcionais, os advérbios hierarquizados na cartografia e pesquisados por Sant'ana (2010) para o PB a partir de Cinque (1999), precisam ser projetados nos Spec XP e marcados com o tipo pré-determinados. Como demonstraremos algumas possibilidades para o PB de um *parser* que segue alguns preceitos gerativista. Utilizamos o processo de geração utilizando o *chart bottom-up* por ser o mais próximo do mecanismo do PM elaborado por Chomsky (1995, 2001).

4.1 MODELAÇÃO E ANÁLISE

A língua portuguesa é uma língua de ordem SVO (Sujeito, verbo e objeto) e se comporta de diversas maneiras na sua ordenação. Os advérbios, a depender de quais os tipos possuem maior mobilidade, podem ter várias posições nesse ordenamento linear. O exemplo abstrato (27) no segundo capítulo demonstra essa ordem categorial dos advérbios na sentença, para cada posição há uma significação para o escopo do advérbio que o qualifica e o identifica. Retomaremos o exemplo (27) abaixo:

(117) (Adv) S (Adv) V auxiliar (Adv) V principal (Adv) O (Adv)

O exemplo acima configura múltiplas regras para os constituintes da categoria advérbio, modelar essas ordens para que se processe tais regras, inicialmente, serve como

introdução à representação dos resultados da fase anterior, onde se projetam as representações linguísticas e extralinguísticas nos sistemas formais. Estabelecer a ordem de acordo com regras internas (AdvP) não favorece bons resultados, tendo em vista que gera dúvidas sobre qual seria o escopo do advérbio por não ser previsto uma projeção de um DP a um AdvP, por exemplo, mas sim, uma adjunção do AdvP, que não é o nosso caso. Diante disso, outras regras deveriam ser registradas para dar conta dessa mobilidade e evitar gerações de sentenças agramaticais.

As evidências de Sant’ana (2010) sobre a posição e geração dos advérbios aplicados no PB são utilizadas na formalização que se segue. As projeções funcionais estabelecidas por Cinque (1999) e, aplicadas no PB, concluem que os núcleos que projetam os advérbios de ato de fala, avaliativos, epistêmicos e evidenciais estariam dentro da camada complementizadora (CP) e, os Atos de fala podem ocupar duas posições no CP. As categorias funcionais se relacionam com as características verbais (Modo, Tempo e Aspecto) notadamente distribuídas na Hierarquia dos Sintagmas Funcionais que determinam o tipo de advérbio e sua função. Uma estrutura que abarque as projeções funcionais para os advérbios devem, inicialmente, ser parte integrante de uma estrutura geral, por isso determinaremos uma estrutura próxima a do PM que possa alocar tais projeções.

É preciso determinar uma estrutura geral e abstrata esperada para as regras gramaticais, partindo da projeção máxima CP até os últimos traços lexicais previstos. Essa necessidade de ter uma estrutura geral é uma imposição do próprio *parser*, por serem aplicações deterministas, precisam ter esquemas formais muito bem determinados. Com isso, estruturamos inicialmente, sem nos deter com detalhes os traços funcionais e lexicais, da definição de um léxico para as regras: o núcleo do sistema implementado. Assumimos que o *parser* atua como um sistema funcional e as categorias lexicais são unificadas com suas propriedades interpretáveis e não interpretáveis. Vejamos abaixo quais as categorias básicas nucleares presentes na estrutura simples que serão representadas de antemão para equilibrar a estrutura.

(118) **C** - Representa dois tipos de sentenças: as declarativas (c) e as interrogativas (QU-).

Seleciona o “T”.

(119) **T** - Representa o núcleo que pode ter traços de tempo e de concordância. Pode receber tanto verbos infinitivo quanto de finitude. Seleciona um “v” e seleciona Caso

nominativo. Com os traços valorados apenas de pessoa, seleciona as propriedades do “v”.

(120) **v** - O “v” representa verbos transitivos, ergativos e inacusativos, diferenciados por informações impressas na MAV.

(121) **V** - Verbos transitivos e inacusativos que selecionam um nome. Verbo inergativos.

(122) **N** - Possui traços formais e Casos. Selecionado por V e spec da projeção máxima de algum núcleo até o nível de “T”.

A árvore que pode ser estruturada para análise de sentenças simples devem acompanhar as especificações acima, não que elas sejam gerais, mas podemos equilibrar e encaixar as projeções máximas para não se limitar apenas na ordem de constituintes. Esses elementos foram vistos anteriormente no exemplo (113) e tal estrutura evidencia uma sentença declarativa em que é explícita os itens funcionais e lexicais selecionados de forma abstrata. Tal estrutura não é suficiente para uma abordagem em que há presença de advérbios, pois não é esperada esse item e a sua projeção, por isso definiremos alguns outros critérios advindos de Sant’ana (2010) para as regras gramaticais que projetarão a estrutura a ser analisada no *parser* a partir do léxico e de algumas sentenças com um tipo de advérbio que comporá o dicionário.

Os advérbios que receberam projeções devem ser motivados graças aos seus valores determinados tais como aponta Cinque (1999), e que estarão nas suas respectivas projeções funcionais acima de TP e abaixo do CP. De acordo com Santos (2011, p. 104), “isso determina a proposta de alguns exemplos de sintagmas adverbiais que ocupem e preencham as posições de núcleos dos especificadores (Spec)”. Tratamos esses advérbios e suas posições de acordo com as projeções funcionais sem a geração de sintagmas funcionais característica da cartografia, apenas informamos em que posição ocupa e o tipo, isso diferencia de certa forma a sintaxe proposta por Cinque (1999), mas não a desvirtua por mantermos as características de tais itens.

Ao manter o traço “T” em vez do IP, referenciado em Cinque (1999), nos justificamos na tentativa de evidenciar os traços de flexão no próprio item com a unificação dos diversos traços morfossintáticos e por tentar se aproximar das questões do PM. Ao observarmos o traço TP e as projeções para os advérbios (<*Mood speech act, Mood evaluative, Mood*

evidential, Mood epistemic, Tpast, Tfuture,... Asp continuative...>). Diante da complexidade em modelar completamente essas funções por criar uma estrutura de regras gramaticais muito detalhista, simplificamos as estruturas para abarcar as projeções correspondentes ao item.

(123) Traços atribuídos ao advérbio “francamente”:

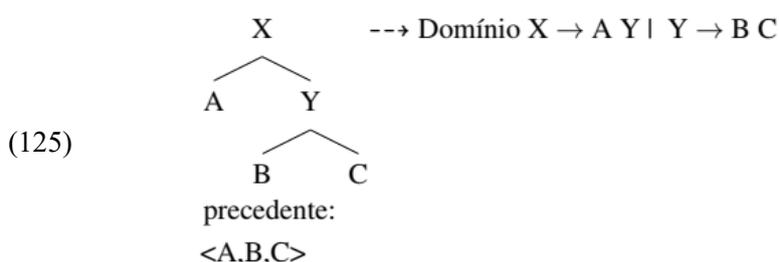
$\left[\begin{array}{l} < \textit{francamente} > \\ \textit{Cat} \\ \textit{Mood} \\ \\ \textit{Pos} \end{array} \right]$	$\left[\begin{array}{l} \textit{Adv} \\ \textit{speech act} \\ \\ \left[\begin{array}{l} < > \\ < \textit{Mood evaluative} > \end{array} \right] \end{array} \right]$
--	---

Os componentes da MAV acima demonstram algumas características básicas do item “francamente”. No atributo Cat(egoria) é definido o valor de Adv(érbio) por ser a classe pertencente do item lexical. O atributo, que na verdade é uma projeção, Mood (Modalidade) é atribuído o valor de *speech act* (Ato de fala) por ser o tipo de projeção pertencente a este advérbio, a sua localização na sentença é definida como um advérbio Alto, o atributo Pos(ição) identifica e localiza a sua ordem hierárquica entre as projeções funcionais, que no caso é “<>” por ser o primeiro valor e não ter nenhum antecessor, não é preenchido, mas o segundo marcador é o <Mood evaluative> (Modalidade avaliativo) que o segue posteriormente. Os Atos de fala estão em uma camada muito acima das demais, alocados em uma camada funcional específica ou no CP. Estes traços e valores formalizados acima se justificam por poderem demonstrar as relações de forma mais explícitas também com as outras categorias que possuem traços (flexão, por exemplo) mais definidas que o advérbio, assim possibilita melhores análises descritivas dos itens lexicais desta classe heterogênea. Esse fator coloca a categoria em evidência, elas já colaboram numa pretensa busca por traços adverbias e as restrições impostas a sua posição. Vejamos abaixo outro advérbio.

(124) Traços atribuídos ao advérbio “provavelmente”:

$\langle \textit{provavelmente} \rangle$ <i>Cat</i> <i>Mod</i> <i>Pos</i>	<i>Adv</i> <i>epistemic</i> $\left[\begin{array}{l} \langle \textit{Mood evidential} \rangle \\ \langle \textit{T(Past)} \rangle \end{array} \right]$
--	--

O fator compensador de não seguir a risca as projeções do modo cartográfico foi inserir o atributo “Pos”, assim visualizamos a ordem esperada nas projeções funcionais. Pode ser considerada opcional este traço “Pos”, devido às ordens e os tipos dos advérbios serem inalterados, ou seja, são de naturezas determinadas nas postulações de Cinque, porém, manteremos devido à exigência de explicitar cada traço para uma melhor descrição sem exigir a atomicidade arbórea que há na cartografia. A organização arbórea deve seguir, além do que já foi demonstrando anteriormente, uma linha mais simples, pois o que nos interessa é uma formalização da posição dos advérbios, tanto em itens precedentes quanto em dominância:



Para Chesi (2004), o precedente é uma realidade fonológica que o *parser* tende a seguir e o domínio é a relação de dominância prevista na X-barras que geram as regras gramaticais, isso se configura como uma limitação que precisa ser contornada, pois o PM não utiliza mais por completo o X-barras, mas sim, um esquema de derivação por fases, por isso, no momento é possível adaptar as regras para quê dê conta das análises sem se distanciar do PM.

(126) Ontem o Sport jogou bem.

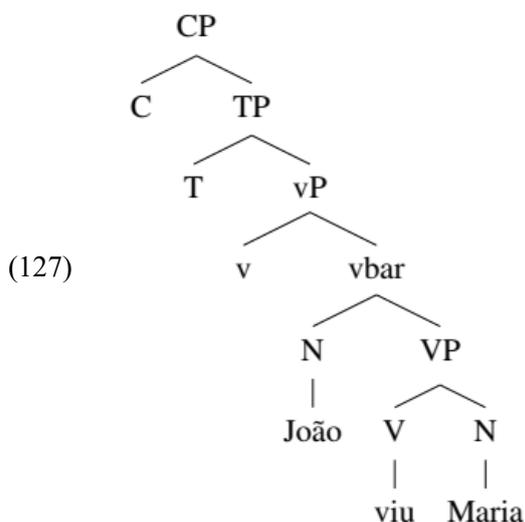
Na sentença acima “Ontem” define o aspecto temporal do verbo “jogou”, entretanto, projetado no início da sentença, as regras que se espera não são formalizadas a contento. O exemplo (126) para Othero (2009b) resultaria em uma estrutura agramatical. Na mesma

perspectiva, para Santos (2009) e David e Othero (2011) tal ordem não é prevista nas suas formalizações, tão pouco é aprofundada essa questão, como já havíamos evidenciado anteriormente.

Outra questão que temos que sanar diz respeito aos precedentes, se estabelecermos regras do tipo:

- (127) $CP \rightarrow C TP$
 $C \rightarrow$
 $TP \rightarrow T vP$
 $T \rightarrow$
 $vP \rightarrow v vbar \mid N vbar$
 $v \rightarrow$
 $vbar \rightarrow N VP$
 $VP \rightarrow V N$
 $N \rightarrow \text{'João'} \mid \text{'Maria'}$
 $V \rightarrow \text{'viu'}$

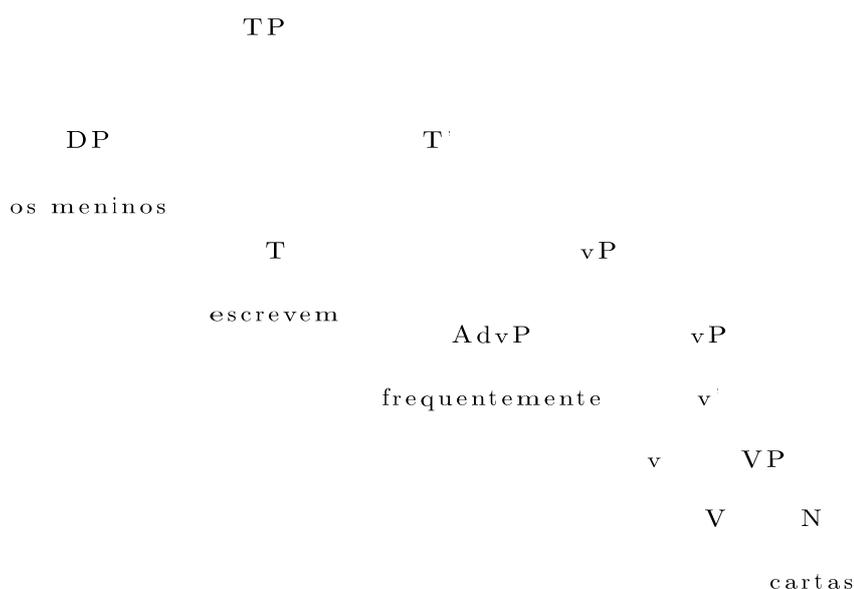
Essas regras estabeleceriam apenas uma organização categorial e não de fato uma análise mais profunda sobre a sentença, tendo em vista que o TP ao c-selecionar o VP, seria projetado visivelmente, e o VP é projetado a partir de V que indicaria o verbo em movimento ou Cópia, como é de se esperar neste tipo de modelo, não seriam produzidos com as regras acima. Abaixo observaremos melhor o resultado de tais regras.



As limitações impostas ficam mais claras quando as regras gramaticais não abarcam as análises que preveem o “João” como spec do TP e “viu” projetado para o T. Mesmo que modificássemos as regras gramaticais, os nós não-terminais V e N seriam preenchidas pelos terminais “João”, “viu” e “Maria”. De outra forma, mas seguindo a estrutura, deveríamos suprimir as categorias base e projetar os itens diretamente nas categorias funcionais, ainda assim não seria possível justamente pelo precedente não permitir inversões da ordem, ou seja, se as categorias estão em ordem linear, o *parser* obrigatoriamente segue essa questão do precedente. Santos (2011) já adianta sobre o problema e determina uma regra que mantém o V vazio, indicando que houve um movimento, assim torna possível uma modelagem melhor, porém, as possibilidades de movimentação dos advérbios torna insuficiente tal regra, como veremos mais adiante. Sendo assim, estabeleceremos uma estrutura que segue a (113) incluindo as projeções funcionais dos advérbios e, os traços unificados devem identificar os traços à que pertence cada item.

Desse modo, ao estabelecermos as posições nucleares individuais na projeção estendida com o advérbio, precisamos determinar que o CP possui uma cadeia para alocar os advérbios Alto de fala acima do TP e as projeções funcionais para os outros advérbios Altos e Baixos estariam no conjunto do TP e vP. No exemplo abaixo retirado e adaptado de Gabriel (2011, p. 175) com o advérbio Baixo “frequentemente” precedido pelo verbo “escrever” verificamos uma das posições habituais dos itens do AdvP:

(127):



Essa posição aponta que o advérbio está precedido pelo verbo, porém, a liberdade que esse item possui nesta sentença colabora na evidência do uso das possibilidades da sintaxe do

PB por este advérbio: (Frequentemente) os meninos (frequentemente) escrevem (frequentemente) cartas (frequentemente). Entretanto, as posições podem indicar algumas mudanças semânticas devido ao seu tipo ser variado, pois temos aqui um caso de advérbio baixo frequentativo (I), mas pode ser identificado como repetitivo (II) em alguma análise do *parser*. A diferença crucial é o tempo do verbo que possibilita identificar o advérbio “frequentemente” como frequentativo (I), basta estar em qualquer tempo verbal, menos o verbo no passado, pois se caracterizará como repetitivo (II). Diante disso é importante estabelecer traços para os outros componentes do dicionário e para as regras nos sintagmas. Abaixo relacionamos algumas sentenças com advérbios para as regras gramaticais.

(130) Francamente, João odeia Maria

(131) Particularmente, acredito em anjos.

(132) Honestamente, João deveria servir o bolo

Para analisar essas sentenças acima é preciso formalizar as regras gramaticais pertinentes. Temos advérbios de Modalidade Ato de fala em posição que indica o tipo de sentença em CP, no caso, é projetado sentenças não assertivas em (131) e (133), porém, é projetado uma sentença assertiva em (132), pois o advérbio em questão “Particularmente”, um Ato de fala, indica que o sujeito do processo verbal é o mesmo da enunciação, projetando uma sentença diferente das outras duas. Partindo então do CP, projetamos fragmento das regras e os itens do dicionário (o léxico):

CP → Spec C'

C' → C TP

C → Adv

TP → DP T'

T' → T VP

VP → DP VP

VP → V DP | V PP

DP → D NP | DP AP | AP DP

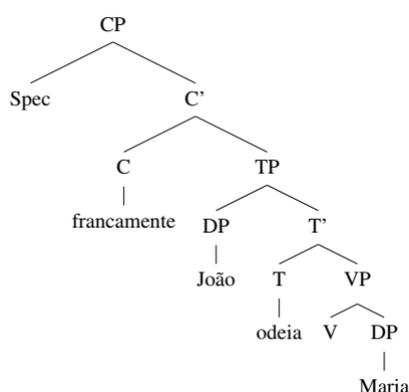
PP → P DP

D → “o”

N → “anjos” | “João” | “Maria” | “bolo”

V → “odeia” | “acredito” | “servir” | “deveria”³⁶

Essas regras possibilitam a análise das sentenças acima, além de outras com a mesma estrutura. Incluímos algumas sentenças lexicais não referenciadas anteriormente, como é o caso do Sintagma Adjetival (AP), O sintagma Preposicional (PP) e o Sintagma Determinante (DP), assim poderemos ampliar mais a análise e evidenciar o posicionamento dos advérbios e seus sintagmas funcionais com mais nitidez. A estrutura esperada para as sentenças possíveis derivadas das regras acima segue abaixo:



Esse fragmento extraído de (SANTOS, 2009), demonstra a ordem dos constituintes sem os traços que favorecem a descrição, contudo, demonstra que há um movimento do verbo que foi projetado do V (vazio) para o T (preenchido com o “odeia”). Ampliaremos as regras e o léxico de acordo com o tipo de sintagma funcional do advérbio presente nas sentenças, assim, as regras ganham contornos para resolver, no mínimo, as questões de ordem e natureza previstas até aqui. Verificaremos como a ordem SVO e as possibilidades de ordenamento do advérbio podem ser formalizadas com mais detalhes. É preciso incluir tais possibilidades de ordenamento sem correr riscos de gerar sentenças agramaticais.

Os valores correspondentes aos outros itens lexicais das demais categorias (N, V, D Adj) são definidas a partir da concordância sintática (Número, pessoa, gênero e tempo) já vistas anteriormente. Esses valores simplificados não serão aqui aprofundados, pois é preciso adentrar nas questões de checagem de traços em relação a esse fenômeno, o importante é configurarmos as categorias para evitar agramaticalidades da seguinte forma: as categorias Nominais (N) possuem traços de número, gênero e pessoa e devem concordar com os traços dos Determinantes (D), as categorias verbais (V) que possuem os traços de número, pessoa e

36. O verbo infinitivo “servir” e o verbo modal “deveria” estão no mesmo nó terminal por questões didáticas, já que possuem alguns traços que os diferenciam, como “tempo” por exemplo.

Essas regras projetam o advérbio correspondente com a sentença, se fixando como escopo ou sobre o TP/VP, a representação acima em que o AdvP representará individualmente os tipos de advérbios correspondentes (AdvP será Adv_actP representando os Atos de fala, por exemplo.). Mantivemos a proximidade com o TP definindo o Advérbio pré-verbais e de início de sentença. As regras que licenciam os pós-verbais são projetados acima ou depois do VP, mas ocorre uma fixação dos advérbios abaixo do TP para visualizar a sua posição, tornando o modelo da estrutura simples, mas como estamos em fase de representação das regras sem realmente estruturarmos a Projeção Funcional (1999, 2006) ou o PM por questões de tempo e espaço, teremos que contornar com regras de ordenamento que explicitam a ordem dos constituintes a natureza do advérbio adotado, como veremos na seção seguinte a atuação das regras, as possibilidades e as limitações encontradas decorrente da formalização rápida das representações.

4.2 O *PARSER* E AS ANÁLISES POSSÍVEIS

Os advérbios em si denotam problemas para qualquer formalização, seja por conta da sua heterogeneidade sintática ou por suas variadas ordens na sentença. Aplicar tal ordem e explicitar a sua natureza demanda tempo e rigorosos testes para a implementação em qualquer sistema de PLN que se projeta sobre a língua(gem) e a esta categoria com itens tão variados. Partindo das pesquisas gerativistas que adotam a projeção até o Spec-XP utilizados como modelo das descrições e análises linguísticas para o desenvolvimento de programas que realizem as análises automáticas, estruturamos as representações com os traços correspondentes, dando ênfase nos núcleos, vejamos abaixo como estruturamos um fragmento as regras gramaticais e o dicionário diretamente no Python-NLTK.

(134):

```

Formalização das Regras

##Complementizador CP##

CP -> Spec Cbar

Cbar -> C Adv_actP | C TP

Spec ->

C ->

##Camada do TP##

```

TP -> Spec Tbar | DP Tbar | Adv_evaluativeP Tbar | Adv_habitualP Tbar | Adv_repetitiveP Tbar

Tbar -> T Adv_evaluativeP | T Adv_habitualP | T Adv_repetitiveP | T vP | T

VP

T -> V

T ->

##Camada funcional vP ##

vP -> Adv_evaluativeP vbar | Spec vbar | Adv_habitualP vbar

vbar -> v VP

v ->

##Camada do VP##

VP -> Spec Vbar

Vbar -> V DP | V PP

V →

As regras em (134) estão distribuídas seguindo o modelo da CFG para uma visualização mais limpa das camadas funcionais. As regras possibilitam gerar algumas sentenças que seguem a estrutura determinada na formalização e algumas outras inesperadas graças as gerações mais espontâneas de algumas sentenças, claro que depende do número de itens disponíveis no dicionário. Para dar conta da ordem dos advérbios, é preciso verificar tanto as projeções antecedentes quanto as posteriores para que não surjam estruturas deformadas ou agramaticais, também tomamos cuidado para não estruturar a versão da X-barra no modelo de P&P devido à utilização da CFG nas projeções funcionais e lexicais.

O Complementizador (CP) possui três regras simples (a regra do nó terminal “C →” atribui um espaço não ocupado por um nó não terminal) que representam a projeção CP, o Spec-CP e o C’ (o C barra, representado como Cbar). É por meio do Complementizador que se identifica uma sentença matriz (declarativa, exclamativa, interrogativa etc.) ou possibilita encaixar uma sentença a outra. Por vezes, é simplificada na sentença, sendo ocultada quando declarativa, assim a sentença é iniciada a partir do TP sem prejuízos a formalização. O fato de incluirmos os advérbios Ato de fala nessa camada é devido as conclusões de Sant’ana (2010). A camada do TP recebe mais regras por ser a projeção fim do verbo no PB, em que a subida para a checagem de traços próprios desta categoria. O destaque está em projetar os advérbios próximos a si mesmo, como o esperado. As demais regras seguem padronizadas para evitar

distorções estruturais, mas, como já adiantamos no início da seção alertando, os advérbios não são padronizados tão facilmente.

Para o dicionário atribuímos uma estrutura de traços com o objetivo de explicitar informações pertinentes tanto morfossintáticas quanto informações advindas da Projeção Funcional, apesar que timidamente, essas informações revelam a natureza da sua projeção e não mais com um simples traço de modificador, como é atribuído em outros modelos gerativos (HPSG e FLG). A vantagem do Python-NLTK em atribuir traços, segundo Bird, Klein e Loper (2014), está no fato de termos liberdade e possibilidades na descrição de um item, de um dado ou fenômeno. Vejamos o fragmento do Dicionário abaixo.

(135)Itens lexicais:

D[CONC= [Num=sg, Gen=masc], Cat = Deter] -> ‘o’

D[CONC= [Num=sg, Gen=fem], Cat = Deter] -> ‘a’

N[CONC = [Num=sg, Gen=masc], Cat = Nome] -> ‘João’ | ‘bolo’

N[CONC = [Num=sg, Gen=fem], Cat = Nome] -> ‘carne’ | ‘porta’ | ‘Maria’

Vaux[CONC = [Num=sg, Per=3], Tense= future, Cat = Verbo] -> ‘deveria’

V[CONC = [Num=sg, Per=3], Tense= pres, Cat = Verbo] -> ‘serve’ | ‘come’ | ‘odeia’ | ‘ama’

V[CONC = [Num=sg, Per=3], Tense= past, Cat = Verbo] -> ‘abriu’ | ‘serviu’

Adv_act[CC= [Cat= Adv], Mood = [‘speech act’], Tipo = [Alto], Pos = [0=‘< >’, B=‘<Mood evaluative>’]] -> ‘francamente,’ | ‘honestamente,’ | ‘realmente,’ | ‘particularmente,’ | ‘absolutamente,’ | ‘felizmente,’ | ‘lamentavelmente,’ | ‘inevitavelmente,’ | ‘surpreendentemente,’ | ‘francamente,’ | ‘sinceramente,’ | ‘acreditadamente,’ | ‘desacreditadamente,’ | ‘verdadeiramente,’ | ‘probabilisticamente,’

Adv_evaluative[CF= [Cat= Adv], Mood = [‘evaluative’], Tipo = [Alto], Pos = [A=‘<Mood speech act>’, C=‘<Mood evidential>’]] -> ‘felizmente’ | ‘inevitavelmente’ | ‘infortunadamente’ | ‘fortunadamente’ | ‘lamentavelmente’ | ‘surpreendentemente’ | ‘curiosamente’ | ‘inesperadamente’

Adv_habitual[CF= [Cat= Adv], Asp = [‘habitual’], Tipo = [Baixo], Pos = [A=‘<Mod possibility>’, C=‘<Asp repetitive (I)>’]] -> ‘usualmente’ | ‘habitualmente’ | ‘costumeiramente’ | ‘regularmente’ | ‘geralmente’

Adv_repetitive[CF= [Cat= Adv], Asp = ['repetitive'], Tipo = [Baixo], Pos =[A='<Asp celerativo (II)>', C='<Asp frequentative (II)>']] -> 'repetidamente' | 'novamente'

No fragmento (135), os itens estão divididos de acordo com a sua categoria e seus traços, a exemplo do D que é dividido de acordo com os traços de gênero e número. Os traços dos itens lexicais são representados da seguinte maneira: [CONC = [Num=sg, Gen=fem], Cat = Nome]. As informações contidas determinam o CONC(ordância) como um atributo e seus valores [Num=sg, Gen=fem], respectivamente número e gênero. O atributo Cat(egoria) tem como valor a própria categoria a que pertence o item. Os traços dos itens advérbios, apresentados em (125), são aqui formalizados da seguinte maneira: Adv_evaluative[CF= [Cat= Adv], Mood = ['evaluative'], Tipo = [Alto], Pos =[A='<Mood speech act>', C='<Mood evidential>']] -> 'felizmente'. O item identificado como Adv é substituído pelo seu tipo funcional, assim a sua explicitação na sentença é mais visível em relação aos demais.

Apesar de serem fragmentos de uma gramática computacional, Essas regras gramaticais e estes itens do dicionário já possuem capacidades de análises. Foram modeladas a partir de sentenças com advérbios de Ato de fala, Avaliativos, Habituais e Repetitivos, respectivamente “Francamente, João serve o bolo.”, “((in)Felizmente) (lamentavelmente) (inevitavelmente) (surpreendentemente) João abriu a porta.”, “João habitualmente come a carne.” e “João (repetidamente) (novamente) come carne”.

Na formulação do fragmento da gramática computacional, utilizamos o programa Geany³⁷ para escrever as regras e o dicionário, e utilizar as ferramentas disponíveis no Python-NLTK na codificação dessas regras de acordo com o modelo utilizado. Diferente de Santos (2009) que utiliza o Donatus (ALENCAR, 2012b), uma ferramenta com interface gráfica que recebe a gramática formalizada em CFG enriquecida por traços e analisa as sentenças de uma forma um pouco mais amigável, a partir do carregamento de um arquivo “.cfg”, nós utilizamos apenas os recursos disponíveis do próprio NLTK e escrevemos os dados no próprio código. Apesar de termos a opção de formalização em separado (arquivo e código) e entendermos ser recomendada esse movimento, os comandos que necessitamos são poucos para a tarefa de análise, por isso consideramos as análises decorrentes do próprio código fonte no momento.

37. ³⁷ Um editor de texto que funciona como um Ambiente Integrado de Desenvolvimento (IDE, do inglês *Integrated Development Environment*). Disponível em <https://www.geany.org/>.

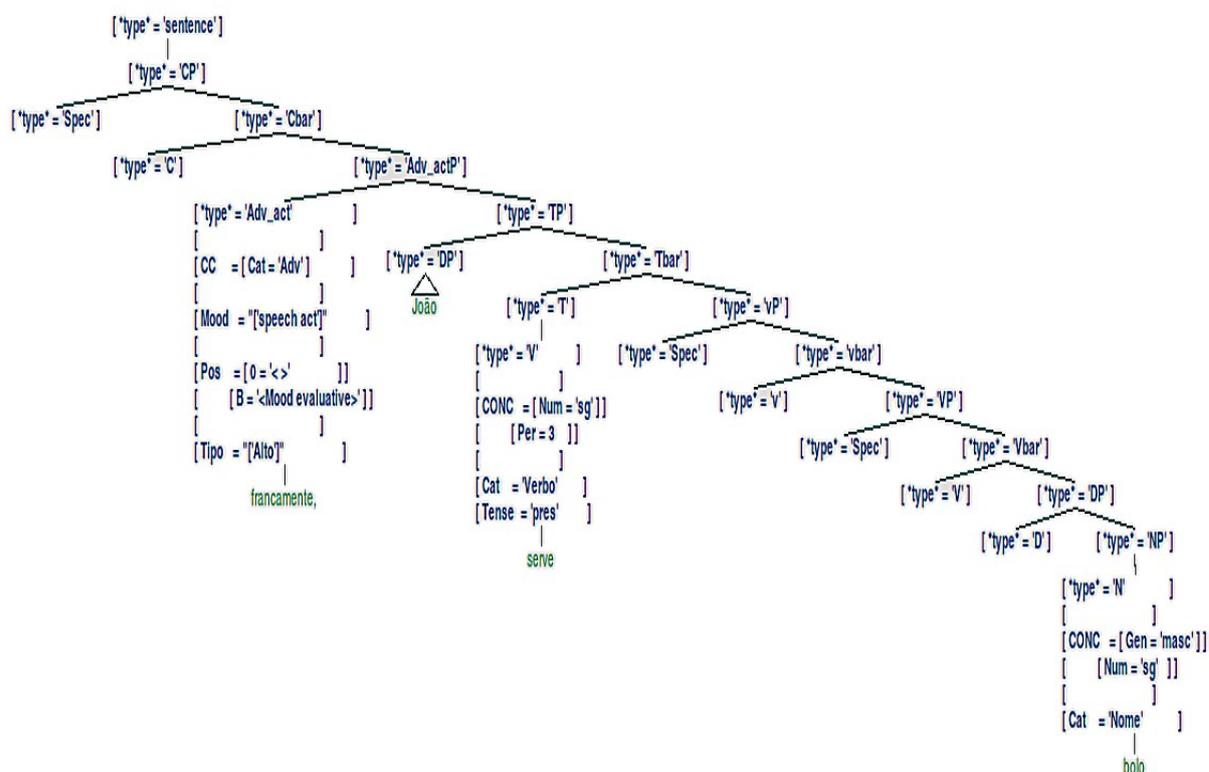


Figura: Representação arbórea da sentença com advérbio Ato de fala “francamente, João serve bolo” possível no NLTK.

A análise demonstra uma sentença em que o advérbio Ato de fala está localizado abaixo da camada do CP e acima do TP; o N “João” se encontra como Spec-TP e o verbo “serve” se projeta no T. As demais projeções que recebem o verbo se encontram vazias, mas indicam o caminho percorrido pelo verbo (com traços no sintagmas funcionais ficaria mais perceptível) e o N “bolo” permanece na base da árvore criada. Não contrariamos Sant’ana (2010), mas decidimos situar o Ato de fala mais próximo ao TP do que dentro do CP. Os traços nos itens lexicais, apesar da organização um pouco diferente do esperado de uma MAV, é possível identificar os traços de concordância, tempo e categoria, confirmando o que Alencar (2009) observou sobre os traços enriquecerem a CFG e torná-la eficiente na formalização computacional de sentenças produtivas.

organizados pelas projeções funcionais básicas do PM e identificando os tipos de advérbios de acordo com a Projeção Funcional, mas sem as múltiplas projeções funcionais que detalham o advérbio por ser extensivo e precisar de projeções vazias até chegar na projeção que está preenchida. Discutimos ao detalhar as regras gramaticais que fariam parte do *parser*, a partir das quais é possível fornecer as previsões de análises que são suportadas pelas regras e, assim, estruturamos os itens lexicais com traços unificados para evidenciar as questões morfosintáticas envolvidas na formação e na ativação da checagem desses valores na estrutura.

Por conseguinte, partimos das possibilidades de análises oferecidas pela linguagem de programação Python e da biblioteca, NLTK, uma ferramenta destinada às pesquisas e estudos em processamento da linguagem natural que visa dentre outros aplicativos envolvendo a linguagem, o uso e criação de *parsers*. Com destaque natural aos advérbios classificados de acordo com os critérios de Cinque (1999, 2006) e aplicados no PB por Sant’ana (2010) e Santos (2011), a listagem dos traços foi facilitada pela ferramenta FCFG³⁸ para identificar cada item e traços.

Com o *parser* estruturado, tendo como referências principais os trabalhos de Santos (2009), Veenstra (1998) e a obra de Alencar e Othero (2011), formalizamos algumas sentenças para servir de modelo para as regras e o léxico presente no dicionário. Licenciemos as regras para o CP, TP, vP, VP e DP e as suas relações na sentença. Modificamos o AdvP, transformando-o de acordo com o seu tipo (Adv_act(P)), assim a análise demonstra com mais nitidez a ordem dos tipos e a natureza. Apesar de carecer de mais informações, as análises que seguem os postulados de Cinque (1999, 2006) e do PM tendem a ter melhores resultados, graças ao poder de formalização e detalhamento, porém, ainda causa vários prejuízos nas tentativas de codificação computacional, visto o exemplo da figura 4, ao se estruturar seguindo as regras, por mais robustas que sejam, os advérbios forçam a utilização de mais projeções, assim como a própria natureza múltipla da categoria, as projeções básicas do PM não são suficientes para a modelagem por exigir duplicação das projeções funcionais, mas não evita a bifurcação e a agramaticalidade.

Nessa perspectiva, as regras que estruturam a sentença “João novamente come carne” utilizou onze linhas de regras para podermos visualizar a estrutura e os movimentos dos itens, dentre todas as possibilidades de formação, a “escolhida” pelo *parser* foi a mais adequada

38. 38 A etiquetagem dos itens são realizadas manualmente, devido à etiquetagem automática possuir um padrão diferente do estabelecido, por exemplo, o “D” é etiquetado como “Det”, e precisaríamos de mais comandos e tratamentos para a etiquetagem dos itens de forma automática, algo que não caberia neste trabalho.

devido as regras gramaticais estruturadas no fragmento da minigramática formada por (135) e (136), apesar de o verbo se fixar na base a colocação do advérbio e a regra prescrita sobre o mesmo evoca mais possibilidades para a reestruturação do fragmento sem quebra estrutural e agramaticalidade das árvores geradas. Para resolver essas questões é preciso mais pesquisas relacionadas ao problema da formalização desses modelos e da própria representação dos advérbios. Devido ao que se tem sobre o PM e sobre a cartografia nos estudos, aqui nos propomos a dar início a resolução dessas questões revelando as possibilidades e limitações da proposta.

Finalizando, mesmo com as limitações, reconhecemos que as pesquisas ligadas aos estudos linguísticos e computacionais são ainda sensíveis ao tocante conhecimento humano da linguagem, tanto em nível de competência, quanto em nível de implementação em aplicativos e programas computacionais que se propõem a utilizar e servir como ferramentas de análises para as pesquisas linguísticas, por tanto, é um desafio em aberto.

5 CONSIDERAÇÕES FINAIS

A busca pela representação linguístico computacional das sentenças com as categorias dos advérbios e codificá-las para o uso em um *parser*, algo ainda não realizado nos estudos em que insere as pesquisas linguísticas e computacionais que tratam da linguagem e do processamento da linguagem natural do PB, para tanto, retomar os estudos e as principais pesquisas sobre a categoria era pertinente para a proposta, porém, as questões sobre os itens assim classificados se mostraram complexas e trouxeram para a representação a problemática que os advérbios por si só carregam. As investigações aqui tratadas não envolveram apenas os aspectos morfossintáticos, mas também as propriedades que licenciam os advérbios na sentença. Algumas dessas informações aqui expostas foram utilizadas na definição de alguns traços unificados e estruturados em matrizes de atributo valor e formalizadas no *parser* com a linguagem de programação Python e a sua biblioteca NLTK.

A dissertação seguiu uma proposta metodológica sobre a aplicação das noções sintáticas a categoria advérbio aplicado no desenvolvimento de um parser em três fases: a fase linguística, a fase linguístico computacional e a fase computacional. Estas fases desenvolvidas por Dias da Silva (1996), determinam as etapas que a pesquisa deve percorrer quando se trata de estudos linguísticos e computacionais na resolução de algum problema ou descrição de algum fenômeno por meio de algum aparato tecnológico.

Na fase linguístico, o objeto por si só é um problema devido a sua natureza e a sua mobilidade. O advérbio é uma categoria que possui diversas características que o distingue das demais categorias que compõem a gramática de uma língua. Diante disto, evidenciar quais as propriedades compartilhadas e a sua relação com as demais classes, além de buscar as razões que definem esta classe heterogênea por ter vários itens tão distintos categorizados como advérbios, retomamos os primeiros estudos a respeito dessa categoria a partir dos gregos helênicos, os mesmos influenciaram as gramáticas posteriores do ocidente e continuam a exercer uma certa influência na tradição gramatical e até nos estudos linguísticos, a exemplo do que se refere nos traços sintáticos (ausência de flexão e variações, além de exercer a função de modificador de outras categorias). Percorremos as classificações dos advérbios nas gramáticas do português utilizadas nas descrições da língua (idealizada) por esses gramáticos, partindo da gramática de Oliveira (1536) até a moderna gramática brasileira de Bechara (2009). Para fazer um contra ponto, utilizamos os estudos advindos da gramática descritiva de Perini (2005), por trazer algumas reflexões sobre o advérbio, tanto na GT quanto em estudos linguísticos mais recentes.

As questões levantadas através dos estudos gerativos espertou espaços para análises e descrições com rigor mais científico. Como vimos Jackendoff (1972) retoma a problemática partindo da mobilidade de acordo com a semântica dos advérbios. Outros autores, como Pollock (1989), descrevem o movimento verbal e observa a influência exercida do advérbio, e Laenzlinger (1998) postula traços que licenciam os itens da categoria e com isso possibilita a movimentação da maioria dos advérbios, em especial, os terminados em *-mente*. Ernst (2004) e Costa (1996) concebem a adjunção do sintagma adverbial na estrutura. Cinque (1999, 2006), contrariando a tradição gerativa que utiliza os advérbios como um indicador do movimento do verbo à IP/TP devido a ideia de os AdvPs serem fixos na estrutura, avança com as questões do Programa Cartográfico e o enriquecimento das estruturas com projeções funcionais que possibilitam o detalhamento do léxico que classifica os advérbios em Altos e baixos, sendo os Altos possuidores de mobilidade maior do que os Baixos que são mais próximos do verbo.

A fase linguístico computacional (representação) e a fase computacional (codificação) demonstraram a importância da formalização e modelação da teoria gerativa para os caminhos na elaboração dos *parsers* que seguem os pressupostos desta teoria linguística. As definições referentes a Linguística Computacional e o contexto histórico das relações entre linguistas e cientistas da computação revelaram entraves entre esses dois tipos de pesquisadores, porém, percebe-se uma união entre essas áreas mais favoráveis atualmente. Demonstramos os tipos de analisadores e como essas ferramentas computacionais estruturam as regras e os componentes do dicionário, a exemplo da CFG. A importância da CFG enriquecidas com traços favoreceram a descrição e a análise das estruturas propostas pela gramática, principalmente pelo poder descritivo que elas possibilitam ao pesquisador, como ao tratar das informações encontradas ou já tipificadas no léxico. A importância da inserção dessas informações sintáticas nos itens na fase linguístico computacional é um requisito na construção e desenvolvimento de léxicos utilizados no processamento natural da linguagem.

Com a utilização da linguagem de programação Python e a sua biblioteca NLTK para uso e manipulação da língua natural em um computador, possibilitou elaborar as ferramentas para o analisador ser capaz de processar sintaticamente os *inputs* (entradas dos possíveis usuários) e assim, formalizar computacionalmente os dados, compondo a terceira fase de forma introdutória. Não se configurou trabalhoso, pois as ferramentas básicas da biblioteca facilitaram, tanto na criação quanto no tipo automático que caracterizou o *parser*.

Contudo, as análises esperadas a partir das informações, até então formalizadas, trouxeram uma série de problemas relacionados a estrutura e a ordem do advérbio na fase

computacional (codificação), pois os mesmos necessitavam de projeções mais detalhadas ou uma organização mais rígida, tal problema era esperado, pois ainda não há por completo analisadores referenciais que estipulem regras formais ao PM, principalmente em sentenças do PB em que ocorrem advérbios. O trabalho de Santos (2009) se propôs a elaborar um parser na perspectiva do PM, mas restringido a visão padrão gerativo sobre essa categoria.

Nossa proposta buscou, primeiramente, por meio de uma perspectiva histórica, levantar os conceitos sobre essa categoria e recorrendo as pesquisas linguísticas de cunho gerativo, buscar mais informações sobre os advérbios, os seus traços e o licenciamento da ordem dos mesmos, com o objetivo de aplicar os resultados obtidos em modelos formais para serem utilizados em um analisador sintático automático que estruturasse a sentença de acordo com o PM e as definições de natureza da categoria advérbio de acordo com os pressupostos cartográficos, tendo como referências as pesquisas de Cinque (1999) e as análises de Sant'ana (2010) aplicados no PB. Diante da falta de uma estrutura rígida formal que modele o PM para ser utilizado em parsers, decidimos limitar a pesquisa e configurá-la como uma introdução aos estudos computacionais da linguagem humana de base gerativa e, continuar a busca dos meios de formalização do PM e do Programa Cartográfico, aprofundando as questões relacionadas ao advérbio em trabalhos futuros. Por isso, o ponto da pesquisa foi demonstrar uma problemática envolvendo a categoria advérbio nas aplicações computacionais e dos estudos que se projetaram sobre a mesma ainda estarem inconclusivos, a sua formalização, que se encontra também em andamento, poderá então abranger os aspectos computacionais e aplicados de forma eficiente nas estruturas que modelem sistemas destinados à competência humana, como é o caso dos parsers que pretendem emular tal capacidade de análise.

REFERÊNCIAS

- ALENCAR, Leonel Figueiredo de. **Teoria da gramática**: uma abordagem computacional. Manuscrito. Fortaleza: [s.n.], 2006.
- _____. Sintaxe formal de línguas não configuracionais num ambiente computacional: o caso do latim. *Revista Virtual de Estudos da Linguagem – ReVEL*, v. 6, n. 10, p. 1–33, 2008. Disponível em: <http://ww.revel.inf.br/files/artigos/revel_10_sintaxe_formal_de_linguas_ao_configuracionais.pdf>. Acesso em: 28 Out 2017.
- _____. Línguas formais, gramáticas e autômatos no processamento automático das palavras. In: ALENCAR, Leonel Figueiredo de; OTHERO, Gabriel de Ávila (Ed.). **Abordagens Computacionais da Teoria da Gramática**. Campinas-SP: Mercado de Letras, 2011. p. 127–152.
- _____. Donatus: uma interface amigável para o estudo da sintaxe formal utilizando a biblioteca em python do nltk. *Alfa*, São Paulo, v. 2, n. 56, p. 523–555, 2012.
- _____. Donatus: uma interface amigável para o estudo da sintaxe formal utilizando a biblioteca em python do nltk. *ALFA: Revista de Linguística*, v. 56, n. 2, 2012. Disponível em: <http://www.scielo.br/cielo.php?script=sci_arttext&pid=S1981-57942012000200008>. Acesso em: 12 Fev de 2018.
- _____; OTHERO, Gabriel de Ávila. **Abordagens Computacionais da Teoria da Gramática**. Campinas-SP: Mercado de Letras, 2011.
- ALLEN, James. **Natural language understanding**. 2. ed. Redwood City: Pearson, 1995.
- AZEVEDO, Jos/’e Carlos de. Sintaxe normativa tradicional. In: OTHERO, Gabriel de Ávila; KENEDY, Eduardo (Ed.). **Sintaxe, sintaxes**: uma introdução. São Paulo: Contexto, 2015. p. 197–216.
- BARBOSA, Jeronymo Soares. **Grammatica philosophica da lingua portugueza**. Lisboa: Academia Real das Sciencias, 1866. Disponível em: <<http://purl.pt/128>>. Acesso em: 21 Ago de 2017.
- BARROS, João de. **Gramática da língua Portuguesa**. Olyssipone: apud Lodovicum Rorigium, Typographum, 1540. Disponível em: <<http://purl.pt/12148>>.
- BECHARA, Evanildo. **Moderna Gramática Portuguesa**. Revista, ampliada e atualizada conforme o novo Acordo Ortográfico. Rio de Janeiro: Nova Fronteira, 2009.
- BEKKER, Immanuel. **Anecdota Graeca**. 1. ed. Berolini: Lexica Segueriana, 1814.
- _____. **Anecdota Graeca**. 1. ed. Graz-Austria: Akademische Druck-u. Verlagsanstalt, 1965.
- BELLETTI, Adriana. On semantic and distributional properties of sentential. **Linguistic Inquiry**, Cambridge: The MIT Press, n. 8, p. 337–351, 1977.
- _____. **Generalized Verb Movement**. Turim: Rosenberg & Sellier, 1990.

BIRD, Steven; KLEIN, Ewan; LOPER, Edward. **Natural language processing with Python: analyzing text with the natural language toolkit**. Sebastopol - CA: O'Reilly Media, Inc, 2014. Disponível em: <<http://www.nltk.org/>>. Acesso em: 11 Dez de 2017.

BOECKX, Cedric. **Linguistic minimalism: Origins, concepts, methods, and aims**. New York: Oxford University Press, 2006.

BOMFIM, Eneida do Rego Monteiro. **Advérbios**. São Paulo: Ática, 1988.

BORGES, Luiz Eduardo. **Python para Desenvolvedores: Aborda python 3.3**. Rio de Janeiro: Novatec, 2014.

BRANCO, Antônio; COSTA, Francisco. **Hpsg: Arquitetura**. In: ALENCAR, Leonel Figueiredo de; OTHERO, Gabriel de Ávila (Ed.). *Abordagens Computacionais da Teoria da Gramática*. Campinas-SP: Mercado de Letras, 2011. p. 219–236.

BRESNAN, Joan; KAPLAN, Ronald M. Lexical-functional grammar: A formal system for grammatical representation. In: BRESNAN, Joan (Ed.). **The mental representation of grammatical relations**. Cambridge: MIT Press, 1982. p. 173–281. Disponível em: <<http://www2.parc.com/isl/groups/nltp/papers/kb82-95.pdf>>. Acesso em: 28 Out 2017.

CÂMARA JR, Joaquim Mattoso. **Estrutura da língua portuguesa**. 30. ed. Petrópolis: Vozes, 1999.

CHESI, Cristiano. Complexity and determinism in linguistic computation: phases and cartography in phrase structure building. *Citeseerx*, Siena, v. 2, p. 32, 2004. Disponível em: <http://www.ciscl.unisi.it/oc/doc_pub/chesi2004-complexity_and_determinism.pdf>. Acesso em: 07 Fev de 2018.

CHOMSKY, Noam. **Syntactic structures**. The Hague: Mouton, 1957.

_____. **Aspects of theory of syntax**. Cambridge: MIT Press, 1965.

_____. **O Conhecimento da Língua: Sua natureza, origem e uso**. Lisboa: Caminho, 1986.

_____. **The minimalist program**. Cambridge: MIT press, 1995.

_____. Derivation by phase. In: KENSTOWICZ, Michael (Ed.). **Ken Hale: A life in language - current studies in linguistics**. Cambridge: MIT Press, 2001. p. 1–52.

CINQUE, Guglielmo. **Adverbs and functional heads: a cross-linguistic perspective**. New York: Oxford Press, 1999.

_____. Issues in adverbial syntax. **Lingua**, Elsevier, v. 114, n. 6, p. 683–710, 2004.

_____. Restructuring and functional structure. In: BELLETTI, Adriana (Ed.). **The Structure of CP and IP: The cartography of syntactic structures volume 2**. New York: Oxford University Press, 2004. p. 132–191.

_____. **Restructuring and Functional Heads: The cartography of syntactic structures**. New York: Oxford University Press, 2006.

COSTA, João. On the behavior of adverbs in sentence-final context. **Linguistic review**, Mouton de Gruyter, v. 14, n. 1, p. 43–68, 1996.

_____. **Subject Positions and Interfaces: The case of european portuguese.** Berlin/New York: Walter de Gruyter, 2004.

DAVID, Karine Alves. **Sintaxe das expressões nominais no português do Brasil: uma abordagem computacional.** 2007. 118 f. Dissertação (Mestrado) — Programa de Pós-graduação em Linguística, Universidade Federal do Ceará –UFC, Ceará, 2007. Disponível em: <http://repositorio.ufc.br/ri/bitstream/riufc/8770/1/2007_dis_kadaavid.pdf>. Acesso em: 04 Nov de 2017.

_____; OTHERO, Gabriel de Ávila. Agrupamentos nominais no português brasileiro: uma análise formal e sua aplicação computacional. In: ALENCAR, Leonel Figueiredo de; OTHERO, Gabriel de Ávila (Ed.). **Abordagens Computacionais da Teoria da Gramática.** Campinas-SP: Mercado de Letras, 2011. p. 77–98.

DI FELIPPO, Ariani. **Representação linguístico-computacional dos adjetivos valenciais do português.** 2004. 120 f. Dissertação (Mestrado em Letras) — Faculdade de Ciências e Letras, Universidade Estadual Paulista, Araraquara, São Paulo, 2004. Disponível em: <<http://www.nilc.icmc.usp.br/nilc/download/Dissertacao.pdf>>. Acesso em: 02 Fev de 2018.

DIAS DA SILVA, Bento Carlos. **A face tecnológica dos estudos da linguagem: o processamento automático das línguas naturais.** 1996. 272 f. Tese (Doutorado) — Faculdade de Ciências e Letras, Universidade Estadual Paulista, Araraquara, 1996.

DUNHAM, Joel Robert William. **The Online Linguistic Database: Software for linguistic fieldwork.** 2014. 419 f. Tese (Doutorado) — The Faculty of Graduate and Postdoctoral Studies Linguistics, The University of British Columbia, Vancouver, 2014.

EMONDS, Joseph. **A transformation approach to syntax.** New York: Academic Press, 1976.

ERNST, Thomas. **The Syntax of Adjuncts.** Cambridge: Cambridge University Press, 2004.

FALK, Yehuda N. **Lexical-functional grammar: an introduction to parallel constraint-based syntax.** Stanford, CA: CSLI Publications, 2001.

FONG, Sandiway. Parsing in the minimalist program: On sov languages and relativization. **The Canadian Journal of Linguistics**, v. 53, p. 237–252, 2008.

FRANCEZ, Nissim; WINTNER, Shuly. **Unification grammars.** Cambridge: Cambridge University Press, 2011.

GABRIEL, Christoph. Questões. In: ALENCAR, Leonel Figueiredo de; OTHERO, Gabriel de Ávila (Ed.). **Abordagens Computacionais da Teoria da Gramática.** Campinas-SP: Mercado de Letras, 2011. p. 153–218.

GONZAGA, Manuela. **Aspectos da Sintaxe Do Advérbio Em Português.** 1997. 215 f. Dissertação (Mestrado) — Faculdade de Letras, Universidade de Lisboa, Lisboa, 1997. Disponível em:

<https://www.academia.edu/777185/Aspectos_da_Sintaxe_dos_Advérbios_em_Português>. Acesso em: 20 Set de 2017.

GUIMARÃES, Maximiliano. Sintaxe minimalista. In: OTHERO, Gabriel de Ávila; KENEDY, Eduardo (Ed.). **Sintaxe, sintaxes: uma introdução.** São Paulo: Contexto, 2015. p. 123–142.

HORNSTEIN, Norbert; NUNES, Jairo; GROHMANN, Kleanthes K. **Understanding minimalism**. [S.l.]: Cambridge University Press, 2006.

JACKENDOFF, Ray S. **Semantic interpretation in generative Grammar**. Cambridge: The MIT Press, 1972.

JEURING, Johan T; SWIERSTRA, S Doaitse. **Grammars and parsing**. Holanda: Utrecht University, 2001. Disponível em: <<http://www.cs.uu.nl/docs/vakken/gont/diktaat.pdf>>. Acesso em: 26 Out 2017.

JURAFSKY, Daniel Saul; MARTIN, James H. **Speech and Language Processing: An introduction to natural language processing, computational linguistics, and speech recognition**. 2. ed. New Jersey: Prentice-Hall, 2010.

KAYNE, Richard. **The antisymmetry of syntax**. Cambridge: MIT Press, 1994.

KENEDY, Eduardo. **Curso Básico de Linguística Gerativa**. São Paulo: Contexto, 2013.

_____. Sintaxe gerativa. In: OTHERO, Gabriel de Ávila; KENEDY, Eduardo (Ed.). **Sintaxe, sintaxes: uma introdução**. São Paulo: Contexto, 2015. p. 11–26.

LADEIRA, Ana Paula. **Processamento de linguagem natural: caracterização da produção científica dos pesquisadores brasileiros**. 2010. 259 f. Tese (Doutorado) — Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Minas Gerais, 2010.

LAENZLINGER, Christopher. **Comparative studies in word order variation: adverbs, pronouns, and clause structure in romance and germanic**. Amsterdam: John Benjamins, 1998.

LIMA, Rafael Bezerra de. **Advérbios focalizadores no português brasileiro**. 2006. 128 f. Dissertação (Mestrado) — Universidade Federal de Alagoas. Faculdade de Letras. Programa de Pós-Graduação em Letras e Linguística, Maceió, 2006.

MAIA, Marcus; FINGER, Ingrid. Introdução. In: MAIA, Marcus; FINGER, Ingrid (Ed.). **Processamento da linguagem**. Porto Alegre: Educat, 2005. p. 9–48.

MCENERY, Anthony Mark. **Computational Linguistics**. Sigma: Sigma Press, 1992.

MIOTO, Carlos; SILVA, Maria C. Figueiredo; LOPES, Ruth. **Novo Manual de Sintaxe**. São Paulo: Contexto, 2013.

MODESTO, Marcello. O programa minimalista em sua primeira versão. In: ALENCAR, Leonel Figueiredo de; OTHERO, Gabriel de Ávila (Ed.). **Abordagens Computacionais da Teoria da Gramática**. Campinas-SP: Mercado de Letras, 2011. p. 127–152.

MONTEIRO, José Lemos. As ideias gramaticais de João de Barros. **Rev. de Letras**, v. 19, n. 1 – 2, p. 25 – 31, 1997.

MORENO FERNÁNDEZ, Francisco. Linguística informática e informática linguística. **Lingüística Española Actual**, v. 12, n. 1, p. 5–16, 1990.

MYKOWIECKA, Agnieszka. Natural-language generation - an overview. **International Journal of Man-Machine Studies**, Elsevier, v. 34, n. 4, p. 497–511, 1991.

NEVES, Maria Helena de Moura. **A vertente grega da gramática tradicional: uma visão do pensamento grego sobre a linguagem**. 2. ed. São Paulo: Editora Unesp, 2005.

_____. O legado grego na terminologia gramatical brasileira. *Alfa*, Araraquara, v. 55, n. 2, p. 641–664, 2011. Disponível em: <<http://seer.fclar.unesp.br/alfa/article/view/4744/4049>>.

OITICICA, José. **Manual de análise: léxica e sintática**. 11. ed. Rio de Janeiro: Francisco Alves, 1955. Disponível em: <<https://archive.org/details/ManualDeAnaliseLexicaESintaticaPorJoseOiticica>>. Acesso em: 12 Ago de 2017.

OLIVEIRA, Antônio Marmo da Cunha. **Dos advérbios**. 1996. 178 f. Dissertação (Mestrado) - Universidade Estadual de Campinas, Instituto de Estudos da Linguagem, Campinas – SP, 1996. Disponível em: <<http://libdigi.unicamp.br/document/?code=vtls000111546>>. Acesso em: 21 Ago de 2017.

OLIVEIRA, Fernão de. **Grammatica da lingoagem portuguesa**. Biblioteca Nacional, 1536. Disponível em: <<http://purl.pt/120>>.

OLIVEIRA, Márcia Santos Duarte de. **Análise sintática do português falado no Brasil**. Rio de Janeiro: Multifoco, 2010.

OTHERO, Gabriel de Avila. **Teoria X-barra: descrição do português e aplicação computacional**. São Paulo: Contexto, 2006.

_____. **A gramática da frase em português: algumas reflexões para a formalização da estrutura frasal em português**. Porto Alegre: EDIPUCRS, 2009. Disponível em: <<http://www.pucrs.br/edipucrs/gramaticadafrase.pdf>>. Acesso em: 17 de Fev 2018.

_____. Um estudo sobre os sintagmas preposicional e adverbial em português brasileiro. In: BORBA, Valquíria C. M.; CARVALHO, Maria Leônia G. Costa; LIMA, Geralda de O. Santos (Ed.). **Contribuições para a pesquisa em linguística nas diferentes áreas: partilhando reflexões e resultados**. Maceió: Edufal, 2009. p. 131–150.

_____; MARTINS, Ronaldo Teixeira. Parsing do português. In: ALENCAR, Leonel Figueiredo de; OTHERO, Gabriel de Ávila (Ed.). **Abordagens Computacionais da Teoria da Gramática**. Campinas-SP: Mercado de Letras, 2011. p. 99–126.

_____; MENUZZI, Sérgio de Moura. **Linguística computacional: teoria e prática**. São Paulo: Parábola Editorial, 2005.

PAGANI, Luiz Arthur. Analisador gramatical em prolog para gramáticas de estrutura sintagmática. *Revel*, v. 2, n. 3, p. 1–27, 2004. Disponível em: <http://www.revel.inf.br/files/artigos/revel_3_analisador_gramatical.pdf>. Acesso em: 4 Nov de 2017.

PEREIRA, Fernando; SHIEBER, Stuart. **Prolog and Natural-Language Analysis**. Stanford: CSLI, 1987.

PEREIRA, Fernando CN; WARREN, David HD. **Definite clause grammars for language analysis – a survey of the formalism and a comparison with augmented transition networks**. Artificial intelligence, Elsevier, v. 13, n. 3, p. 231–278, 1980.

PERINI, Mário A. **Gramática descritiva do português**. 8. ed. São Paulo: Ática, 2005.

PINTO, Luciano César Garcia. **Do que se confia às letras: A ciência gramatical nas etimologias de isidoro de sevilha**. 2008. 420 f. Dissertação (Mestrado) — Universidade Estadual de Campinas – Unicamp, Instituto de Estudos da Linguagem, Campinas, 2008.

POLLARD, Carl Jesse; SAG, Ivan A. **Head-driven phrase structure grammar**. Stanford, Calif: CSLI Publications, 1994.

POLLOCK, Jean-Yves. Verb movement, universal grammar, and the structure of IP. **Linguistic Inquiry**, Cambridge: The MIT Press, v. 20, n. 3, p. 365 – 424, 1989. Disponível em: <<http://www.jstor.org/stable/4178634>>. Acesso em: 21 Set de 2017.

RIBEIRO, Julio. **Grammatica portugueza**. 2 Ed. refundida e muito augmentada. São Paulo: Teixeira & irmão, Editores, 1885. Disponível em: <<http://www2.senado.gov.br/bdsf/handle/id/518640>>.

RIZZI, Luigi. **Relativized minimality**. Cambridge: The MIT Press, 1990.

ROBOREDO, Amaro de. *Methodo grammatical para todas as linguas (1619) – facsimilada*. prefácio e estudo introdutório de carlos assunção e gonçalo fernandes. Vila real: Centro de estudos em letras, universidade de trás-os-montes e alto douro: Coleção Linguística 1, 2007.

RODRIGUES, Érica dos S; AUGUSTO, Marina R. A. Modelos formais de gramática: o programa minimalista vs. gramáticas baseadas em restrições – hpsg e lfg. **Matraga**, Rio de janeiro, v. 16, n. 24, p. 133–149, 2009. Disponível em: <<http://www.e-publicacoes.uerj.br/index.php/matraga/article/view/27800/19921>>. Acesso em: 21 Out de 2017.

SAID ALI, Manoel. **Gramática elementar da língua portuguesa**. 9. ed. São Paulo: Edições Melhoramentos, 1965.

_____. **Gramática histórica da língua portuguesa**. 19. ed. São Paulo: Livraria Acadêmica, 1971.

SAKSIDA, Maria José Agostini. **Lições acerca do advérbio: Uma viagem diacrônica por gramáticas do português**. 2005. 150 f. Dissertação (Mestrado) — Pontifícia Universidade de Católica de Minas Gerais, Programa de Pós-graduação em Letras, Belo Horizonte -MG, 2005.

SANT'ANA, Mauro Simões de. **Sintaxe e Processamento de Advérbios no Português Brasileiro**. 2010. 150 f. Tese (Doutorado) — Faculdade de Letras, Universidade Federal do Rio de Janeiro - UFRJ, Rio de Janeiro, 2010.

SANTOS, Andréa Feitosa dos. **Movimento do verbo e categoria vazias em I e em V em um fragmento de gramática computacional do português**. 2009. 96 f. Dissertação (Mestrado) - Pós-Graduação em Linguística, Universidade Federal do Ceará, Fortaleza -CE, 2009. Disponível em: <<http://www.repositorio.ufc.br/handle/riufc/2818>>. Acesso em: 11 Dez de 2017.

SANTOS, Paulo Roberto Pereira. **Os sintagmas adverbiais predicativos de constituintes no português brasileiro: uma perspectiva cartográfica do IP**. 2011. 180 f. Dissertação (Mestrado) — Instituto de Letras, Universidade Federal da Bahia – UFBA, Salvador, 2011. Disponível em: <<http://repositorio.ufba.br/ri/handle/ri/8617>>. Acesso em: 1 de Out 2017.

SARDINHA, Tony Berber. A língua portuguesa no computador. In: SARDINHA, Tony Berber (org). **Ver a língua portuguesa no Computador**. Campinas-SP: Fapesp, 2005. p. 7–32.

SERVILLE, Isidore of. **Etymologies**. Intro. Trad.: Stephen A. Barney et al. Cambridge: Cambridge, 2006.

SHIEBER, Stuart M. Evidence against the context-freeness of natural language. *Linguistics and Philosophy*, **Springer**, v. 8, n. 3, p. 333–343, 1985.

SILVA, Débora Maria Pereira da. **Aquisição da camada flexional e um estudo exploratório da relação entre os advérbios e as categorias tempo e aspecto**. 2015. 266 f. Dissertação (Mestrado) - Programa de Pós-Graduação em Letras, Pontifícia Universidade Católica de Minas Gerais, Belo Horizonte, 2015.

SILVA, Gláucia V. **Word order in Brazilian Portuguese**. New York: Mouton de Gruyter, 2001.

TEIXEIRA, Zenaide Dias. **Propriedades Sintáticas e Semânticas dos Advérbios no Português Brasileiro**. 2015. 173 f. Tese (Doutorado) — Universidade de Brasília – UNB, Brasília, 2015.

VEENSTRA, Mettina Jolanda Arnoldina. **Formalizing the minimalist program**. 1998. 199 f. Tese (Doutorado) — University of Groningen, Holanda, 1998. Disponível em: <<https://www.rug.nl/research/portal/files/10654819/thesis.pdf>>. Acesso em: 12 Out de 2017.

VIEIRA, Renata; LIMA, Vera Lúcia Strube de. **Linguística computacional: princípios e aplicações**. IX **Escola de Informática da SBC-Sul**, Passo Fundo, Maringá, São José: SBC-Sul, v. 3, p. 47–86, 2001. Disponível em: <<http://www.inf.unioeste.br>>. Acesso em: 4 Nov de 2017.